VOLUME 18, 1956/57, No. 2 pp. 33-72 Published 27th July 1956

Philips Technical Review

DEALING WITH TECHNICAL PROBLEMS
RELATING TO THE PRODUCTS, PROCESSES AND INVESTIGATIONS OF
THE PHILIPS INDUSTRIES

EDITED BY THE RESEARCH LABORATORY OF N.V. PHILIPS' GLOEILAMPENFABRIEKEN, EINDHOVEN, NETHERLANDS

PHASE LINEARITY OF TELEVISION RECEIVERS

by A. van WEEL.

621.375.121.018.782.3:621.397.62

The picture signal that controls the beam current of the picture tube in a television receiver must be a faithful copy, both in amplitude and in phase, of the signal generated in the television camera. The properties thus required of the vision channel of receivers could be met fairly simply while there were comparatively few television transmitters in existence, i.e. while receivers were not required to be particularly selective. The situation has changed, however, since transmitters became more numerous, and the selectivity of receivers had to be improved accordingly. In this connection, difficulties have arisen with regard to the phase linearity. In some quarters recourse has been made to compensating defective phase linearity at the transmitting end. The author of this article puts forward another solution, that of a phase-linear I.F. amplifier, which turns out to be no more expensive than that of conventional I.F. amplifiers.

It should be noted that the work described in this article is based upon the Gerber standard (so named after the Chairman of the C.C.I.R. sub-committee which proposed this standard) now in use in a number of countries. This standard prescribes negative picture modulation, and frequency modulation for the sound. In these respects it is similar to the Amercian standard, but different from the British standard. As a consequence, the author's conclusions are directly applicable to the American system, but only with certain restrictions to the British system.

The "vision channel" of a television receiver embraces all the circuits from the aerial terminals to the picture tube. The aerial delivers a carrier that is amplitude-modulated with the picture signal. The vision channel has the task of producing a signal exactly corresponding to that modulation for application to the control electrode of the picture tube. A block diagram of the vision channel appears in *fig. 1*.

The picture signal is a complicated function of time, being proportional to the brightness values of the various picture elements scanned one after the other in the camera. The networks of the vision channel must be designed so as to reproduce this time function undistorted.

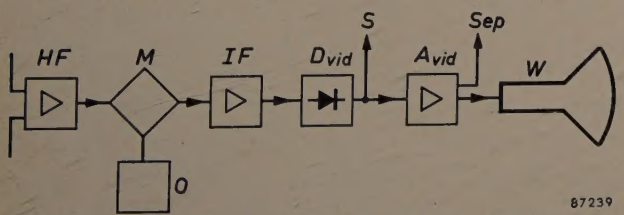


Fig. 1. Block diagram of the vision channel in a television receiver. *HF* high frequency amplifier. *O* local oscillator. *M* mixer stage. *IF* intermediate frequency amplifier. D_{vid} video detector. A_{vid} video amplifier. *W* picture tube. *Sep* sync. separation stage. In the intercarrier sound system, the sound signal is taken off at *S*.

It is possible to compute the shape of the output signal from any circuit, given the shape of the input signal. A *step function* (*fig. 2a*) is frequently chosen as input signal; the amount of distortion exhibited by the output signal (fig. 2*b*) gives a good idea of the effect of the network on the picture signal it passes [1]).

In practical television engineering, however, it is not usual to employ this method for individual sections of the circuit. In preference, use is made of Fourier analysis whereby any time function can be regarded as the sum of a large number of sine functions. If the time function is not periodic the Fourier spectrum consists, not of a number of discrete lines (as does that for a periodic time function), but of a continuous band. Thus, with a non-periodic time function, the power contained in the original is distributed over a certain range of frequency.

[1]) Cf. J. Haantjes, Philips tech. Rev. **6**, 193-201, 1941.

The reason for preferring this method of approach to that of step functions is that it is usually easier to calculate how a network transfers single sinusoidal components at different frequencies than to investigate directly how the network in question would react on the time function itself.

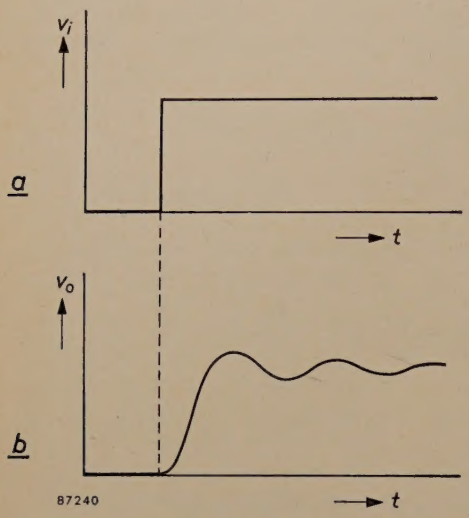


Fig. 2. *a*) Step function and (*b*) typical step function response. v_i: input voltage, and v_o: output voltage of a network, plotted as function of time t.

Knowing the frequency range within which the Fourier components of the signal fall, the networks are required to pass this frequency range without distortion. This means that the intensity relationship between the various components may not be changed (a change of strength of all components in the same ratio is of course permissible). Likewise, the individual components may not undergo differing delays (again, all components may be delayed to the same extent, since this merely means delaying the signal as a whole). Furthermore, no new components such as noise, hum and non-linear distortion products may be added; in the present article, however, we shall not consider these matters.

What is required, then, of all sections of the circuit along which the signal has to pass is that both their amplitude characteristics and their delay-time characteristics shall be flat for the frequency range in question.

A sinusoidal component of the input signal may be written $\hat{v} \sin \omega t$; at the output this becomes $a\hat{v} \sin \omega(t-\tau_f)$. The quantity a, plotted as a function of frequency, gives the *amplitude characteristic;* the quantity τ_f similarly plotted gives what is called the *phase-delay characteristic* [2]). The latter represents the retardation or delay undergone by a given phase of the component in question. The retardation may also be interpreted as the phase angle $\varphi = \omega\tau_f$ that arises on account of the network. To make the phase delay independent of frequency, the phase-angle must increase linearly with frequency. It follows from this that for the undistorted transmission of a signal all the networks employed must be phase-linear.

Relationship between the amplitude and delay characteristics

It is worth while to consider the causes of the delay in signal transmission just discussed. In line transmission it is easy to see that the finite velocity of propagation of the waves along the cable involves a certain delay. In networks consisting of coils, capacitors and resistors the physical distance to be travelled by the signal is very short; that marked delays nevertheless occur in them is a consequence of the fact that a certain time is necessary for both the building up of an electric field in a capacitor and the building up of a magnetic field in a coil. Only in a perfect resistor are the voltage and the current exactly in phase.

It is impossible to build amplifying circuits that, apart from tubes, contain only ohmic resistances, for resistors necessarily involve a certain amount of stray capacitance, and so do tubes. The effect of these capacitances can be reduced by employing lower resistance values, but then the amplification falls off as well.

It is possible to obtain high and constant amplification within the desired frequency band with the aid of special networks ("wide-band circuits"). The wider the band and the higher the amplification required, the greater is the number of coils and capacitors needed in these networks. The inertia of the network therefore also increases. In general, this additional delay will not be the same for all frequencies, so that the delay characteristic will vary all the more as the effectiveness of the network is improved from the point of view of amplitude alone. If on the other hand the network is so designed that the delay properties are improved, then there is a deterioration in the amplitude response.

The demand for as much amplification as possible within a certain frequency band thus creates difficulties with regard to the amplitude characteristic or the delay characteristic or to both; over and above this, the need to satisfy certain selectivity requirements outside the desired band also leads to undesired effects within the band. Thus, under the Gerber Television Standard (CCIR), the video channel must exhibit an amplitude characteristic which is flat up to about 5 Mc/s and which falls to 5% at

[2]) H. J. de Boer and A. van Weel, An instrument for measuring group delay, Philips tech. Rev. **15**, 307-316, 1953/54.

5.5 Mc/s (under this standard the frequency difference between sound and picture carriers is 5.5 Mc/s). All that is required of the delay characteristic is that it should be flat up to 5 Mc/s; the delay of suppressed components is of no importance in television receivers.

The amplitude characteristic described above can always be obtained with the aid of sufficiently complicated filters. To make the amplitude characteristic run flat up to a given cut-off frequency and then fall very steeply, certain resonance effects are made to occur in the neighbourhood of the cut-off frequency. Now, resonance phenomena have the property of taking a certain time to come into action; passive resonant networks can naturally supply no energy of themselves and it is only possible to obtain high amplitudes of oscillation by a building-up process. The inference is that the phase delay of such networks is always markedly greater in the region of the cut-off frequency than in the remaining part of the frequency band.

In general, minimum phase-shift networks are employed for this purpose. These are networks which at all frequencies have a smaller phase-shift than that obtainable with other networks giving the required amplitude characteristic. The question of whether a given network is of minimum phase-shift type or not, may be resolved by the possibility or otherwise of drawing an equivalent circuit of the network in the form of a cascade of π or T sections, or both, using only realizable circuit elements (see *fig. 3a*). If this is impossible on account of bridging elements (fig. 3*b*), then the network is, in general, not one of a minimum phase-shift type.

It must be borne in mind that the criterion may only be applied once the network has been reduced to its simplest form, should this be necessary. If, for example, Z_1, Z_2, Z_3 in fig. 3*b* are impedances of the same kind, a π-T transformation makes it possible to reduce the whole circuit to a single T-section made up of realizable elements (fig. 3*c*). Hence the network is in fact of a minimum phase-shift type.

In the practical design of filters that have to exhibit a linear phase characteristic within the bandpass range, a common procedure is first to give the amplitude characteristic the desired shape with the aid of minimum phase-shift networks, and then to add a special kind of non-minimum phase-shift network that comes within the category of "all-pass" networks. These have the property of combining a horizontal amplitude characteristic for all frequencies from zero to infinity with a phase delay which in one way or another varies with frequency; they thus offer an opportunity of compensating the phase errors of the first filter. Such circuits are accordingly to be found in professional television equipment, e.g. in rado relay systems, in amplifiers for coaxial cables and, as will be discussed later, in certain television transmitters.

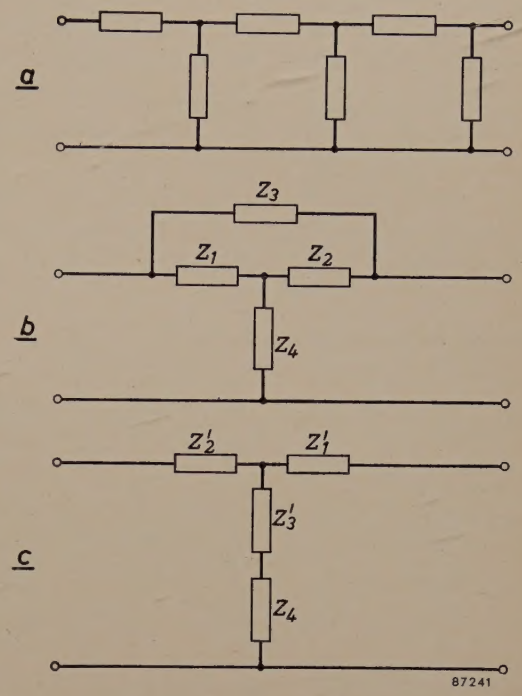


Fig. 3. *a*) Example of a network consisting of a T-section and π-section in cascade. If the network is made up entirely of realizable circuit elements, then it is of a minimum phase-shift type.
b) If Z_1, Z_2 and Z_3 are impedances not of the same kind, the network is not, in general, of minimum phase-shift type. If they are of the same kind, the delta arrangement can be turned into a star arrangement, as in (*c*), so that network (*b*) is in fact of minimum phase-shift type (provided that all the impedances are realizable ones).

The use of such phase-compensating networks in normal television receivers is impossible for economic reasons. However, experience has shown convincingly that there is too much variation in the phase delays in modern receivers. These errors are much more troublesome in recent sets than they were in older ones — a direct consequence of the growth in the number of transmitters, which has made it necessary to increase receiver selectivity. The increased selectivity has been achieved by the employment of circuits that fulfil that purpose but at the same time involve greater phase-delay errors.

The selectivity of a television receiver, like that of a normal broadcast receiver, is obtained in the IF amplifier. The general rule that higher selectivity means bigger discrepancies in the phase delays is valid here as elsewhere. In the IF amplifier there is amplification of an IF carrier modulated with the video signal. The absolute value of the delay under-

gone by the carrier is in no way important for the quality of the picture received; the necessary condition is that the carrier and the side-band components should all have the same delay. As was explained in the article referred to above [2]), it is not necessary in such a case that the phase delay φ/ω of the IF amplifier should be constant over the pass-band; it is sufficient that the *group delay* $\tau_g = d\varphi/d\omega$ should satisfy this requirement because, when it does so, the phase delay of the modulation is the same for all video frequencies.

The Gerber Standard recommends an amplitude characteristic for the receiver as well as one for the transmitter (*fig. 4*). It will be seen from fig. 4*b*

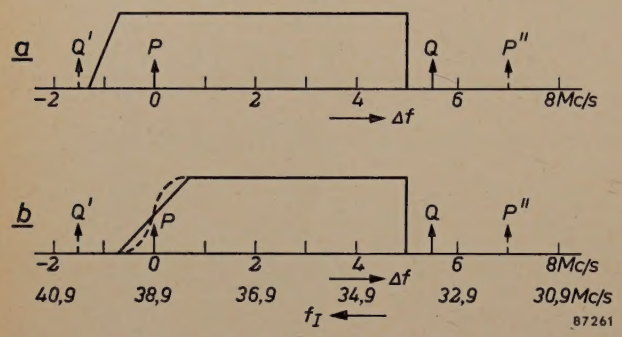


Fig. 4. *a*) Frequency spectrum of a transmitter according to the Gerber Standard of the C.C.I.R. (tolerances omitted). The horizontal scale represents difference of frequency Δf from the picture carrier. *P* indicates the picture carrier and *Q* the sound carrier. *P''* is the picture carrier of the next channel above, and *Q'* the sound carrier of the next channel below. *b*) The corresponding amplitude characteristic for the receiver. The scale of intermediate frequencies f_I is applicable to the (usual) choice of 38.9 Mc/s as IF for the picture carrier.

that at the receiving end only one side-band has to be dealt with in respect of modulation frequencies of more than 0.75 Mc/s. The phase-delay τ_{fm} of the modulation in this single side-band is given by:

$$\tau_{fm} = \frac{1}{\omega_m} \int_{\omega_0}^{\omega_0+\omega_m} \tau_g \, d\omega, \quad \ldots\ldots \quad (1)$$

where ω_0 is the angular frequency of the carrier and ω_m the angular frequency of the modulation.

For modulation frequencies under 0.75 Mc/s there is asymmetrical double side-band reception. Here the phase delay of the modulation follows, in a somewhat more complicated fashion, from the ratio of the amplitudes of the two side-bands as well as from the magnitude of τ_{fm} as given by equation (1) for each side-band separately (the limits of integration for the lower side-band being ω_0 and $\omega_0 - \omega_m$; see also the article referred to in [2])). Here too it remains true that, if the group delay is constant within the IF band to be amplified, then the phase delay of the modulation is constant for all video frequencies; on the other hand, deviations from a flat group delay characteristic generally give rise to discrepancies in the phase delay of the modulation frequencies.

The IF amplifier must therefore possess a flat group delay characteristic over the pass-band. The amplitude characteristic within this range — and outside it too (for reasons of selectivity) — must fulfil certain requirements. It is not necessary for the amplitude characteristic of the receiver to embody a slanting straight flank as shown in fig. 4*b* or an approximation to it. Any characteristic that is centrally symmetrical about the carrier frequency (like the dotted curve in fig. 4*b*) provides a flat amplitude curve for the modulation.

Effect of the amplitude characteristic on the shape of the signal

At the highest modulation frequencies the amplitude characteristic is rectangular in shape (fig. 4*b*). It is not *a priori* certain that cutting off the picture frequency band gives the best quality picture. To explain this in more detail, let us consider an imaginary network having an exactly rectangular amplitude characteristic that cuts off abruptly at the frequency $f_{lim} = \omega_{lim}/2\pi$ (*fig. 5a*), and giving a constant delay-time for all frequencies; for the sake of simplicity we shall suppose the delay-time to be zero. If now such a network is fed with a signal having the form of a step function (fig. 5*b*) it can be fairly easily shown that the output signal will be given by the integral sine function:

$$v_0(t) = \frac{2}{\pi} \int_0^{\omega_{lim}} \frac{\sin \omega t}{\omega} \, d\omega. \quad \ldots \quad (2)$$

The shape of this function of t can be worked out by expanding it as a series; fig. 5*c* shows the result. The following three properties of the function will be noted:

(1) The output voltage deviates from zero even before $t = 0$ (and even when $t = -\infty$);

(2) v_0 is not infinitely steep at $t = 0$;

(3) On both sides of $t = 0$ the curve goes beyond the initial and final values; it oscillates about them and approaches them as the distance from $t = 0$ increases.

That the function should "begin" for negative values of t, i.e. even before a voltage is present on the input, is of course physically impossible. The result obtained is a consequence of the assumption that the delay-time of a filter of this kind is zero, and this too is physically unobtainable. A filter

with an amplitude characteristic like fig. 5*a* can be approached as closely as desired if one is prepared to extend it with more and more sections. A perfectly rectangular amplitude characteristic will

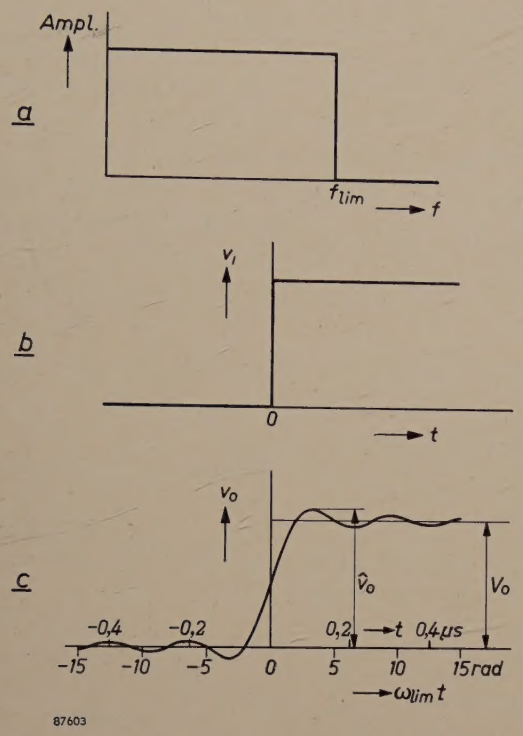


Fig. 5. *a*) Flat amplitude characteristic of an imaginary network, falling abruptly to zero at the cut-off frequency f_{lim}. *b*) Step function (cf. fig. 2*a*).
c) Response to (*b*) of network with amplitude characteristic as in (*a*). The overshoot factor is defined as $(\hat{v}_0 - V_0)/V_0$.

only be obtained when an infinite number of sections is employed. The filter will then have an infinitely long delay-time; in other words, the instant $t = -\infty$ at the output according to fig. 5*c* is the same as the instant $t = 0$ at the input according to fig. 5*b*. This explains the apparent irreality of point (1) above.

It may be added that it is possible to arrive at a fairly close approximation of the amplitude characteristic in fig. 5*a* using only a limited number of elements. In such a case, phase errors will occur and these will exactly compensate the oscillations in fig. 5*c* during the times prior to zero; here too, therefore, no output signal will arise before the input signal is applied.

The second property mentioned above, the finite slope at $t = 0$, is directly connected with the bandwidth available. The abscissa in fig. 5*c* is given in units of $\omega_{lim}t$; the greater $\omega_{lim}\,t$, the steeper will be the rise in the curve when plotted as a function of t. Thus the bandwidth determines the steepness obtainable.

The third property, the occurence of "undershoot" and "overshoot" — as the fluctuations around the initial and final values may be termed — is directly connected with the amplitude characteristic's abrupt fall to zero at the cut-off frequency. It can be shown [3]) that these oscillations do not occur when the amplitude characteristic, instead of being rectangular, is Gaussian in form, i.e.

$$A = A_0 \exp(-a\omega^2), \quad \ldots\ldots \quad (3)$$

as in *fig. 6a*, curve *I*. If the parameter a in (3) is chosen such that the Gaussian curve has already dropped considerably when the cut-off frequency f_{lim} of fig. 5*a* is reached, it will be seen that the reproduction of the step function will be much less steep (see fig. 6*b*). A suitable compromise between steepness and amount of undershoot and overshoot is obtained by giving the amplitude characteristic the form of a Gaussian curve up to the cut-off frequency, where it will have fallen to one half, and then letting it drop to zero abruptly (curve *II* in fig. 6*a*). The initial oscillation then has an overshoot factor of about 5%, as against 9% for the integral sine function.

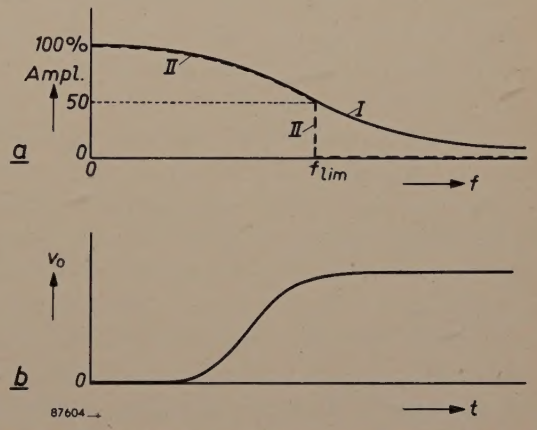


Fig. 6. *a*) *I* amplitude characteristic of Gaussian form. *II* the same, except that the characteristic falls abruptly from 50% to zero at the cut-off frequency f_{lim}.
b) Step function response of a network having the amplitude characteristic *I* in (*a*).

Selectivity requirements and phase errors

So far we have only considered the amplitude characteristic for the modulation signal. We shall now lay down in greater detail the selectivity required outside the admitted band. Firstly, the sound of the desired channel must be suppressed (the carrier frequency of the sound signal is separated from the picture carrier frequency by 5.5 Mc/s);

[3]) See, for example, C. Cherry, Pulses and transients in communication circuits, Dover Publications, New York 1950, p. 175, or K. Küpfmüller, Die Systemtheorie der elektrischen Nachtrichtenübertragung, S. Hirzel, Zurich 1952, p. 51.

secondly, the risk of interference from transmitters adjacent in frequency must be considered. Since a band of 7 Mc/s is available for each television channel the picture carrier is separated from the sound carrier of an adjacent channel by only 1.5 Mc/s (Q'-P in fig. 4). The sound signal from a transmitter working on the adjacent channel may cause interference when the receiver is tuned to a weak transmitter. In these circumstances the signal from the adjacent sound carrier penetrates as far as the video detector and sets up a video frequency component of 1.5 Mc/s, which can produce very troublesome effects in the picture.

For these reasons it is necessary that the neighbouring sound carrier should be heavily attenuated; in good receivers the standard is that it should be 1% or less of the maximum of the IF amplitude curve (100-fold attenuation). On the other side of the pass-band, the sound carrier of the desired channel has to be suppressed. This implies 200-fold attenuation, but it need not be effected entirely in the IF amplifier; the corresponding video frequency (5.5 Mc/s) lies outside the picture frequency band, and the sound of the desired channel can therefore be suppressed partly in the video amplifier. In receivers using the intercarrier sound system [4]), 20-fold attenuation (to 5%) is often laid down for the IF section.

None of this is particularly difficult to put into practice as long as only the amplitude characteristic need be considered. But complications proceed from the effect mentioned above, whereby discrepancies in delay-time arise near frequencies corresponding to abrupt changes in the amplitude characteristic. This is illustrated by the curves in *figs. 7a* and *b;* the former is the selectivity curve [5]) and the latter the group delay curve of the IF amplifier of a normal television receiver. If a comparison is made between the curves from the middle of the band outwards, it will be seen that on both sides the group delay starts to change before the selectivity. It seems as if coming events with regard to the amplitude response cast their shadows before them in the group delay behaviour.

The effect of group delay changes in the IF section on the modulation signal is much more serious at the picture carrier end (i.e. for low modulation frequencies) than it is at the other end of the admitted band (high modulation frequencies). This may be inferred from a calculation of the phase delays of different modulation components by means of equation (1). In this equation the group delay is integrated over a certain frequency band and then divided by the modulation frequency in question. Two hypothetical group delay curves are drawn

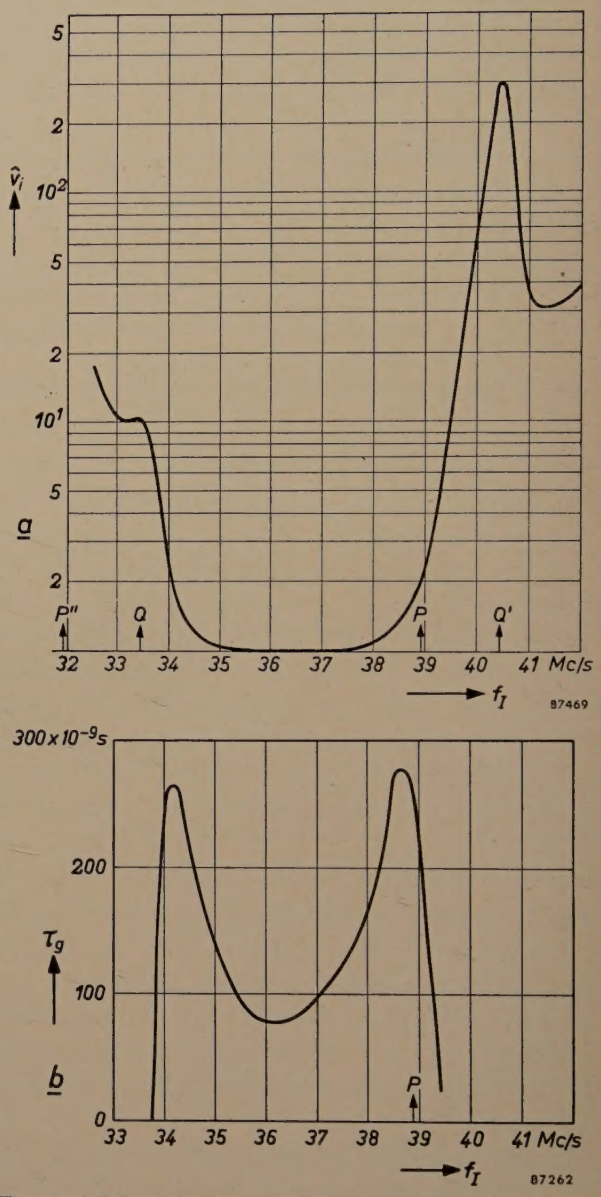


Fig. 7. *a*) Selectivity curve and (*b*) group delay characteristic of the IF amplifier of a normal television receiver. (In (*a*) the input signal $\hat{v}_i$ required to give a certain output signal is plotted on a logarithmic scale.) With regard to P, Q, P'' and Q', see fig. 4.

in *fig. 8a*: curve *I* exhibits a change in τ_g in the vicinity of the carrier frequency but is otherwise flat; curve *II* exhibits a change of the same magnitude, but at a point well away from the carrier frequency. It will be seen that the integral of equation (1) has the same value for $\omega_m = \omega_{m1}$ in curve *I* as for $\omega_m = \omega_{m2} \gg \omega_{m1}$ in curve *II* (the two hatched areas in the diagram are equal); but to find the values of τ_{fm} we still have to divide

[4]) See for exemple Philips tech. Rev. **15**, 198, 1954/55.

[5]) The selectivity curve is the curve obtained by plotting as ordinates the reciprocals of the ordinates of the amplitude curve.

the integral by ω_{m1} and ω_{m2} respectively. Since ω_{m1} is smaller than ω_{m2}, the group delay characteristic represented by curve *I* has a greater interfering effect on the picture than that represented by curve *II*.

In fig. 8*b* the same effect is demonstrated in the group delay curve measured on an actual IF amplifier. In this case the carrier frequency was 38.9 Mc/s. The phase delay (τ_f) of the video frequency signal (insofar as it originates in this side-band) was determined by graphical integration from the group delay characteristic as measured. It may be seen that the big changes in τ_g at high video frequencies cause only limited changes in the phase delay.

Thus the phase errors arising from the incorporation of trap circuits for the adjacent sound channel — which, as indicated by fig. 4, lies only 1.5 Mc/s away — are much more troublesome than the phase errors that occur at higher modulation frequencies (up to 5 Mc/s) owing, for example, to the suppression of the desired sound carrier (which lies 5.5 Mc/s away). In practice, the phase errors of the first-mentioned, more troublesome, variety manifest themselves mainly as a smudging of abrupt black-to-white transitions in a left-to-right direction.

Fig. 9 illustrates the control voltage response at the picture tube for two step functions, one a step from white to black and the other a step from black to white. The full line represents the case in which the phase distortion under discussion is absent, the broken line a case where it is present. The bulge into blacker-than-black preceding the second step is, of course, not visible on the screen; it can be seen, however, that the jump in brightness is not from black to white but from black to grey producing a gradual, smudged, transition towards white. The received picture suffers a considerable loss of definition in consequence.

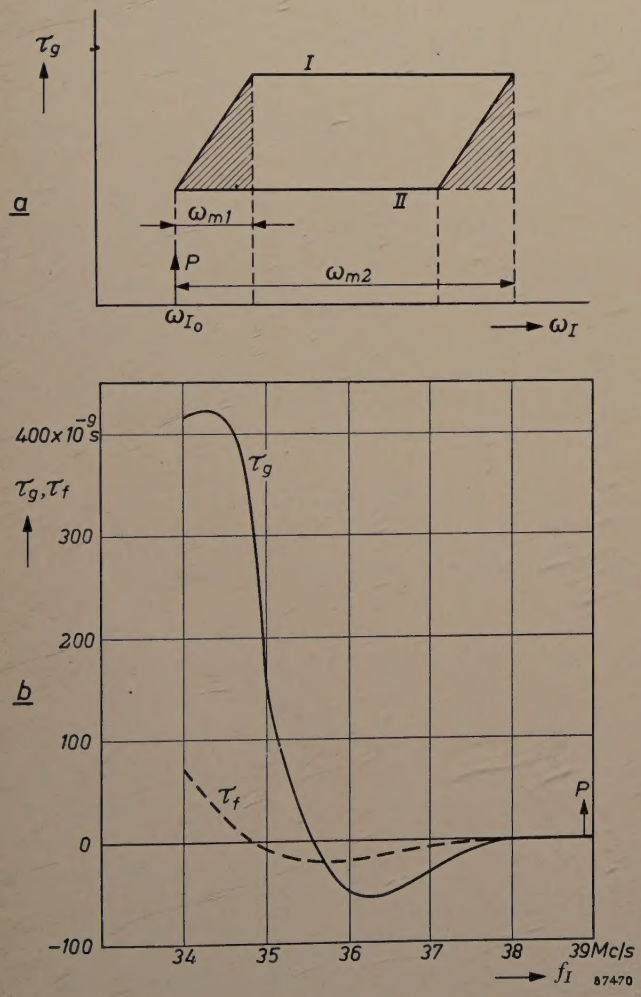


Fig. 8. *a*) Comparison of two hypothetical group delay characteristics (*I* and *II*) for an IF amplifier. *II* is better than *I*, since its rising portion is further away from *P*, the IF of the carrier.
b) Measured group delay characteristic τ_g of a certain IF amplifier. The phase delay time τ_f has been derived from τ_g by graphical integration.

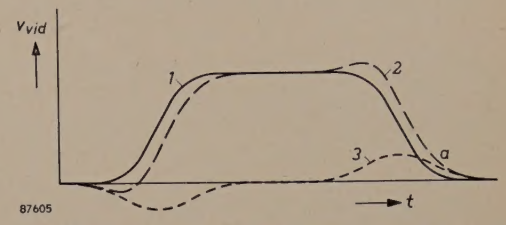


Fig. 9. Response to a white-to-black transition, followed by that to a black-to-white transition. v_{vid} control voltage on the picture tube. The full black line *1* represents the case where phase distortion is absent, the broken line *2* (which is the sum of curves *1* and *3*) the case where it is present. The divergence of *2* from *1* at *a* (on the extreme right) manifests itself in the picture as a grey smear.

Possible solutions

There are three ways in which the unfavourable effects of phase errors might be removed:
(1) by compensation in the video section of the receiver,
(2) by compensation in the video section of the transmitter, and
(3) by the use in the receiver of an inherently phase-linear IF amplifier.

The last-named method seems at first to be the most obvious, but up to now it has not been used. This is a consequence of a curious situation which has developed in the field of television receiver design. It is generally known that phase properties are of importance in television; nevertheless, the phase characteristic as such has not hitherto been taken into consideration, measured or checked either in the design or the manufacture of television sets. This applies to almost all the monochrome receivers at present in existence.

Two reasons can be advanced to explain this strange state of affairs. For one thing, the measurement of phase delay characteristics has become a subject of study only since the war. Various equipments for this purpose have meanwhile been developed, but in general they have been so complicated as to be not suitable for employment in development laboratories, to say nothing of employment in the factory. It is only quite recently, at the Philips Research Laboratories in Eindhoven, that apparatus in the early days of a new technical development.

It may be added that one kind of check on phase properties is always possible, viz. the judgement of picture quality; further, an impression of the receiver's behaviour in this respect may be gained by displaying on an oscillograph the response of the whole receiver to a step function. In these two ways it is quite feasible to decide whether distortion is present, but to pronounce on the kind and the cause of the distortion is another matter.

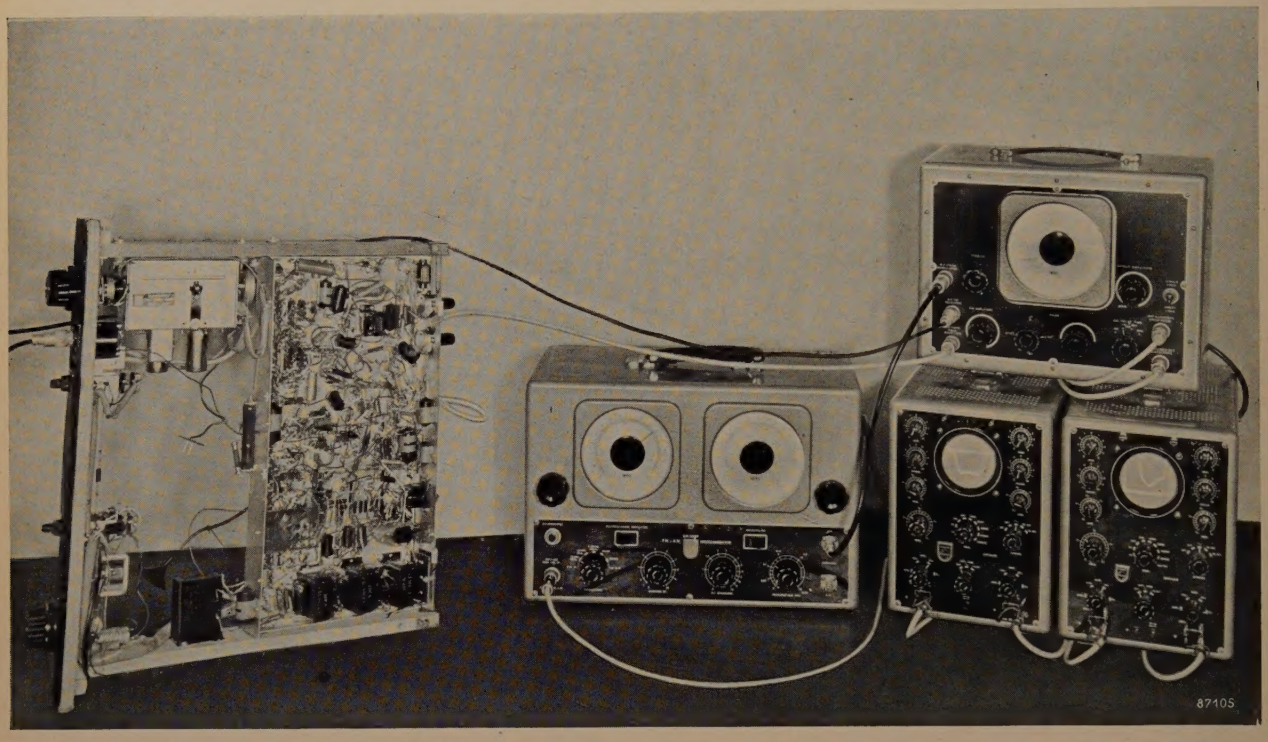


Fig. 10. Checking the selectivity and group delay characteristics of the IF amplifier in a television receiver.
Left: Receiver chassis. *Centre:* Wobbulator (Type GM 2889 signal generator). *Right:* Two oscillographs and, resting upon them, a device by means of which the selectivity curve and the group delay characteristic are displayed simultaneously on the left-hand and right-hand oscillographs respectively. This group delay meter has been in production for some time, under the type code GM 2894.

(*fig. 10*) has been developed[6]) which allows the group delay characteristic to be displayed directly on the screen of a cathode-ray tube, and which is so simple that it can be used for checking in manufacture.

Thus the direct measurement of phase properties was a difficult matter until recently. The design of television receivers was therefore necessarily based on the amplitude characteristic alone. In the beginning results were fairly good. This can be attributed to the fact that in those days it was not necessary to make the standard of selectivity as high as it is today. Apart from this, it is, in general, not usual to aim extremely high

In the meantime the falling-off in picture quality due to improved selectivity was not watched idly in development laboratories. Means to overcome it were sought in an empirical fashion. Giving the amplitude characteristic of the VF amplifier a pronounced hump between 1 and 2 Mc/s — until then, of course, it had been made as flat as possible — seemed to provide a remedy, up to a point. It did indeed effect a considerable improvement in picture quality. Thorough investigation of the processes associated with this modification of the VF amplitude characteristic has shown that it created phase errors in the video amplifier that almost cancelled out those in the IF amplifier. Amplitude distortion was also involved, it is true, but its effects

[6]) C. J. Heuvelman and A. van Weel, Wireless Engr. **33**, 107-113, May 1956.

proved to be less troublesome than those of the original phase erors. Thus picture quality was improved, without however becoming optimum for the given bandwidth.

Given the carrier frequency and the amplitude and the group delay characteristics of an IF amplifier, it is possible to calculate accurately the phase distortion undergone by the video frequency signal. The phase errors can then, in principle, be compensated to any desired extent in the video frequency section with the aid of the all-pass networks mentioned above. This is not done in practice, for it would make the receiver too expensive. It is therefore a case of making do with the the imperfect but inexpensive solution described above.

A further solution — point (2) above — is to use video frequency phase compensation at the transmitter. At the transmitting end the question of cost is of much less importance, and in no way prohibits the use of this method. One disadvantage that suggests itself is that, for this solution to be universally acceptable, all television receivers would have to have the same IF group delay characteristic; this is of course not so in practice. There are other disadvantages associated with this solution; we shall return to them at the end of the article.

It is perhaps useful to devote a few words to the phase properties of the transmitter. The filter that has to suppress the unwanted side-band in the antenna signal likewise causes considerable phase errors; it is not practicable to compensate them in the radio-frequency section on account of the high power that has to be fed to the antenna. Consequently the compensation is always carried out in the video frequency section of the transmitter, where it must balance the phase errors due to the vestigial filter as well as possible.

The effect of the compensation is checked by observing the picture appearing on the screen of a receiver, say, of a standard commercial type. If the compensation is adjusted in such a way that picture quality is optimum, then the phase errors of the receiver being used are compensated at the same time. All those who possess a set of this type will therefore be able to receive a good quality picture, but the quality of picture given by receivers with better phase properties will be worse!

For this reason, compensation at the transmitter of phase faults in the reveiver must be based on a standardized receiver delay-time curve. The Nordwestdeutsche Rundfunk has worked out this problem systematically. The basis chosen was a receiver IF amplitude characteristic which fulfilled certain demands as to selectivity and which was capable of realisation with the aid of minimum phase-shift networks. On this basis it is possible to work out the corresponding group delay curve, and upon this the compensation in all transmitters and the design of future receivers has to be based. This solution, as adopted by the Nordwestdeutsche Rundfunk, is doubtless one of the best thought-out practical forms of the system of transmitter compensation.

The reasons that recourse was made to phase compensation at the transmitting end was the assumption that a selective receiver could only be made phase-linear by the employment of complicated and therefore expensive circuits. In what follows we shall show that this belief is unjustified and the need for phase compensation at the transmitter is then seen in quite a different light.

Requirements to be fulfilled by a phase-linear IF amplifier

Before going into the ways and means of realizing a phase-linear IF amplifier, we shall set down the requirements to be fulfilled as regards the amplitude characteristic.

Inside the pass-band, the amplitude characteristic must fall by a factor of two at the IF carrier frequency, and elsewhere possess the desired shape from the standpoint of the video frequency modulation.

Outside the pass-band it must satisfy the following demands:

(1) At the frequency of the sound carrier, the amplification must fall to approximately 5% (it may not be lower with the system of intercarrier sound [4]) in general use in the CCIR system).

(2) At the frequency corresponding to the sound carrier of the *adjacent* channel the amplification must fall to 1% or less.

(3) Beyond this frequency the suppression may be less severe, since the modulation depth of the highest picture frequencies in the neighbouring picture channel is always fairly slight. Under the Gerber Standard (C.C.I.R) the ratio of the powers of sound and picture carrier is 1 : 5. The sound carrier may be regarded as a single side-band modulation of the picture carrier with a frequency of 5.5 Mc/s. With 100% depth of modulation, the power of each side-band component would be a quarter of the carrier power. With a sound power/picture power ratio of 1 : 5, as above, the depth of modulation of the sound carrier, regarded as side-band component, is $\sqrt{4/5} \times 100 = 89\%$. If it is enough to attenuate this "modulation component" to 1%, then an attenuation to 5% will suffice for the much weaker picture-modulation components of the highest frequencies.

(4) For the two adjacent picture carriers, lying at +7 and —7 Mc/s from the desired picture carrier, the signal must likewise be attenuated to 1% or less; this can always be achieved without difficulty.

Information as to how these demands on the amplitude curve are to be met and how at the same time a linear phase characteristic within the pass-band may be obtained, is provided by the fundamen-

tal relationship between the phase and amplitude characteristics of networks with minimum phase shift. In an earlier section it was mentioned that any change in the shape of the amplitude characteristic necessarily affects the shape of the phase or delay-time characteristic. For networks with minimum phase-shift it is possible to give an analytical relationship connecting the amplitude and phase characteristics such that the one curve is wholly determined when the other is known for all frequencies from zero to infinity. This relationship, known as Bode's relation [7]), can be written as follows:

$$\tau_{fc} = \frac{\omega_c}{\varphi_c} = \frac{2}{\pi} \int_0^\infty \frac{A - A_c}{\omega^2 - \omega_c^2} \, d\omega, \quad . \quad . \quad . \quad (4)$$

where τ_{fc} is the phase delay at the angular frequency ω_c, φ_c the phase shift at this same frequency and A the "insertion loss" (a logarithmic measure of the attenuation in the network concerned, i.e. the amplification in the network is proportional to e^{-A}).

For the time being we shall disregard the ideal amplitude characteristic and attempt to work out from equation (4) what sort of amplitude characteristic results from the requirement that the phase characteristic should be linear. The frequency band for immediate consideration will be that of the video frequencies, that is, from zero to a certain cut-off frequency.

For the phase characteristic to be linear, the phase delay must be constant; that is, the value of the integral in (4) must not depend on ω_c. This is so if A is a quadratic function of ω:

$$A = a\omega^2, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

since the quantity to be integrated then reduces to a. However, this gives an infinite delay. It is not necessary to seek a physical explanation for this, for the application of equation (4) to an amplitude characteristic in accordance with (5) is not permissible, the reason being that certain limitations must be introduced in deriving (4) and these limitations exclude a curve in accordance with (5) [8]).

Nevertheless, it is possible to use the quadratic-exponential amplitude curve given by (5) as a basis for further investigation, owing to the effect of the weighting factor $1/(\omega^2 - \omega_c^2)$ in the integral of (4). This weighting factor is very large in the immediate vicinity of ω_c. This being so, the phase angle and phase delay at these frequencies — although of course depending on the values possessed by A throughout the whole frequency range from zero to infinity — will be determined largely by the way that A changes in the neighbourhood of ω_c.

This property allows the following statement to be formulated: *if it is ensured that the attenuation factor A is a quadratic curve up to a certain cut-off frequency and that outside this frequency band it does not diverge too rapidly from this quadratic form, then the phase characteristic may be expected to be linear within the said frequency band, with significant deviations occurring only in the vicinity of the cut-off frequency.* This expectation is confirmed in practice, as we shall see later.

Thus our first conclusion is that the amplitude characteristic of a phase-linear video frequency amplifier within the pass-band is given by $\exp(-a\omega^2)$, that is, that the characteristic is a Gaussian curve.

However, we were concerned not with video frequency amplifiers but with IF amplifiers, i.e. with band-pass networks. Now there is a network theorem which says that any video frequency characteristic can be practically realised in the form of half the characteristic of a bandpass filter (see *fig. 11*) providing that the band it passes is relatively narrow.

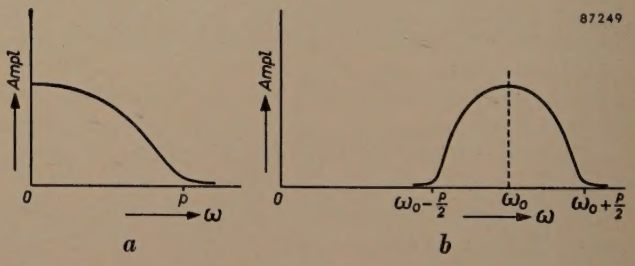


Fig. 11. *a*) Amplitude characteristic of a low-pass network over the range $\omega = 0$ to $\omega = p$.
b) Amplitude characteristic of a band-pass network over the range $\omega = \omega_0 - \frac{1}{2}p$ to $\omega = \omega_0 + \frac{1}{2}p$.
Curve (*a*) can be regarded as the right-hand half of curve (*b*) which has been shifted through a distance of ω_0 to the left, the abscissa being expanded by a factor of 2.

By virtue of this principle, the amplitude and phase properties discussed above can be transferred directly to the band passed by a filter. Thus, to give a bandpass filter a linear phase characteristic, the selectivity characteristic must take the form of a Gaussian curve within the pass-band. IF selectivity curves are usually drawn with a logarithmic amplitude scale; the Gaussian curve is then a parabola.

[7]) H. W. Bode, Network analysis and feedback amplifier design, Van Nostrand, New York 1953, pp. 302-336.
[8]) J. Peters, Einschwingvorgänge, Gegenkopplung, Stabilität; J. Springer, Berlin 1954, pp. 43-44.

Before going into the question of whether a curve of this kind is suitable for the transmission of a picture signal modulating a carrier wave, we shall show how far a Gaussian amplitude curve is compatible in practice with a linear phase characteristic or, in this case, with a flat group delay characteristic.

In wide-band amplification it is usual to work with staggered circuits; this consists of employing *LC* elements that are slightly detuned and differently damped with respect to one another in the various stages of a cascade amplifier. The *LC* elements are so designed that the amplitude characteristic of the whole amplifier is as flat as possible.

It is also possible to design the elements so that the group delay is as nearly constant as possible. In *figs 12* and *13* amplitude and group delay curves for one half only of the pass-band are drawn for an amplifier containing various numbers of *LC* elements. In fig. 12 the aim was constant amplitude, in fig. 13 constant group delay. It will be seen from fig. 12 that, as the number of circuits n increases, the amplitude curve remains flat up to an ever higher cut-off frequency, in the vicinity of which, however, the group delay curve deviates more and more from flatness. In fig. 13 it is seen that, the greater the number of *LC* elements, the greater is the range of frequency over which there is an almost constant group delay, while the amplitude characteristic within this range exhibits ever greater divergences.

In order to show to what degree the amplitude curves of fig. 13 approximate to a Gaussian function, we have transferred them to the logarithmic scale

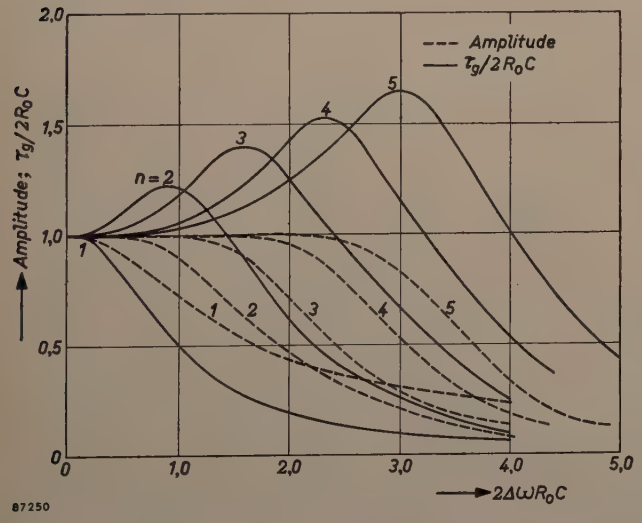


Fig. 12. Amplitude and group delay curves over one half of the band admitted by a wide-band amplifier with n staggered circuits, for various values of n. The quantity R_0C is a measure of the average impedance of the n circuits. The aim in view was to attain as flat an amplitude curve as possible.

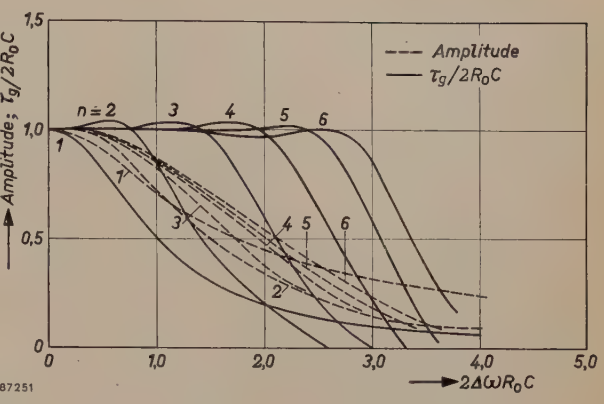


Fig. 13. The same as fig. 12, except that the aim in view was as constant a group delay as possible.

of *fig. 14*, in which the crosses denote the case with three *LC* elements and the circles the case with six elements. The full lines drawn in the figure are parabolae, the parameters of which are chosen such that the curves agree as closely as possible with the transferred amplitude curves for low abcissa values. With three *LC* elements, marked divergence from the parabola begins at an abscissa value of 1.6; with six *LC* elements it begins at an abscissa value of 2.8. We see from fig. 13 that these are indeed the values at which the group delay curves begin to deviate considerably from flatness.

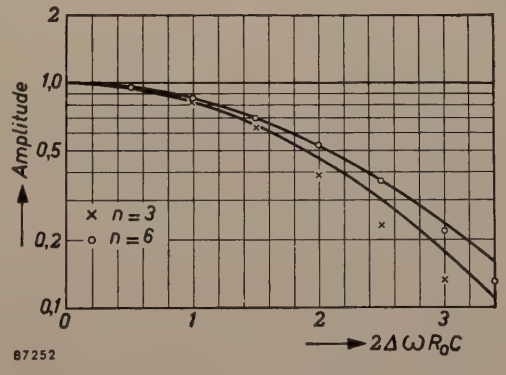


Fig. 14. Crosses: Points of the amplitude characteristic for $n = 3$ in fig. 12. Circles: the same for $n = 6$. The full curves are the parabolae that most nearly fit the crosses and circles at low absissa values.

The Gaussian curve as an IF selectivity curve

In order to investigate the suitability of a Gaussian function as amplitude characteristic for an IF amplifier, we have drawn such a function on a logarithmic scale (the parabola I_a) in *fig. 15a*. The parameter of the parabola has been chosen such that the amplification at the frequencies both of the IF picture carrier (38.9 Mc/s) and of the side-band component corresponding to a modulation frequency of 4.5 Mc/s (i.e. 34.4 Mc/s) falls to one half. It is an easy matter (if the modulation depth

is small) to work out the corresponding amplitude curve for the modulation by reading the values of the two side-band components of each modulation frequency from curve I_a, and adding each pair. The result is curve I_b in fig. 15*b*.

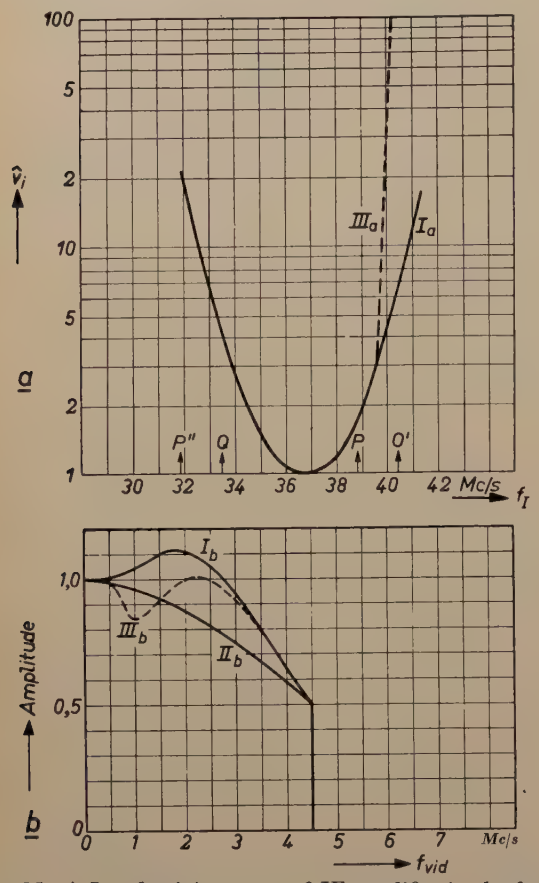


Fig. 15. *a*) I_a: selectivity curve of IF amplifier in the form of a parabola (Gaussian function). III_a: selectivity required for suppression of the adjacent sound carrier Q'.
b) Modulation amplitude characteristics. I_b corresponds to I_a in (*a*), III_b to III_a. The desired characteristic is II_b (cf. curve II in fig. 6*a*).

Curve II_b may be taken as an ideal modulation-amplitude curve; it falls a little on approaching the cut-off frequency, to reduce the overshoot factor. It will be seen that there is not a great deal of difference between curves I_b and II_b; if the selectivity it increased (as is required), the disparity becomes even less significant. Disregarding for a moment the possible effect on the delay curve, consider the curve III_a that does satisfy the selectivity requirement; we see that the corresponding modulation-amplitude curve III_b approaches the ideal II_b even more closely. However, we shall not attempt further improvement of the amplitude curve by building on these somewhat hypothetical foundations, but return instead to the subject of phase properties.

The problem now is how to modify the Gaussian selectivity curve in such a way that selectivity requirements are met while the group delay curve is nevertheless kept flat in the band-pass range. This can be done to quite a close approximation by making use of Bode's relation (eq. (4)), according to which the phase at a given frequency is mainly determined by the shape of the amplitude curve in the immediate vicinity of that frequency. Let us write A, the attenuation factor in (4), as

$$A = A_1 + A_2,$$

where $A_1 = a(\omega - \omega_c)^2$, the amplitude variation is accordance with curve I_a in fig. 15, and A_2 is the additional attenuation required to bring about the desired suppression of the adjacent sound carrier. The equation for the phase delay then becomes:

$$\frac{\varphi_c}{\omega_c} = \frac{2}{\pi} \int_0^\infty \frac{A_1 - A_c}{\omega^2 - \omega_c^2} \, d\omega + \frac{2}{\pi} \int_0^\infty \frac{A_2}{\omega^2 - \omega_c^2} \, d\omega. \quad (6)$$

Now A_1 was a Gaussian function, and consequently the first term on the right-hand side of (6) is again independent of ω_c; thus it is a matter of choosing A_2 such that the second term has only a small effect in the neighbourhood of the picture carrier frequency. One solution of the problem is to make A_2 zero throughout except at those frequencies where additional suppression is necessary, as illustrated in *fig. 16*. The magnitude of A_2 at these frequencies is determined by the attenuation required; the smaller the band is over which A_2 is not zero, the less will be the effect on phase at the picture carrier frequency, for the value of the second integral in (6) will then be smaller.

Thus, to start with, we may stipulate that suppression must be confined to the narrowest possible band of frequencies, in other words that highly selective trap circuits must be employed.

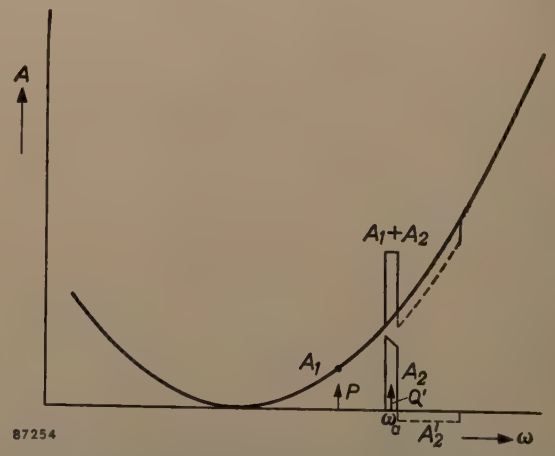


Fig. 16. A_1 attenuation factor of parabolic form, i.e. a Gaussian function. A_2 the additional attenuation necessary for suppressing the adjacent sound carrier Q'.

A second qualitative requirement can be deduced from this general theory. If we add a negative portion to function A_2, as shown by a dotted line in fig. 16, the total effect of the second term on the right-hand side of (6) will be further reduced for frequencies around the carrier frequency. Experiments have in fact shown that circuits of which the amplitude curve is re-entrant beyond the suppressed frequencies do possess a group delay curve that is flatter over the pass-band. Clearly it is not desirable to push the selectivity in this frequency region beyond the point of real necessity.

Design of a phase-linear IF amplifier

The foregoing considerations have been taken as the basis for the design of an IF amplifier (see *fig. 17*) that satisfactorily fulfils the requirements.

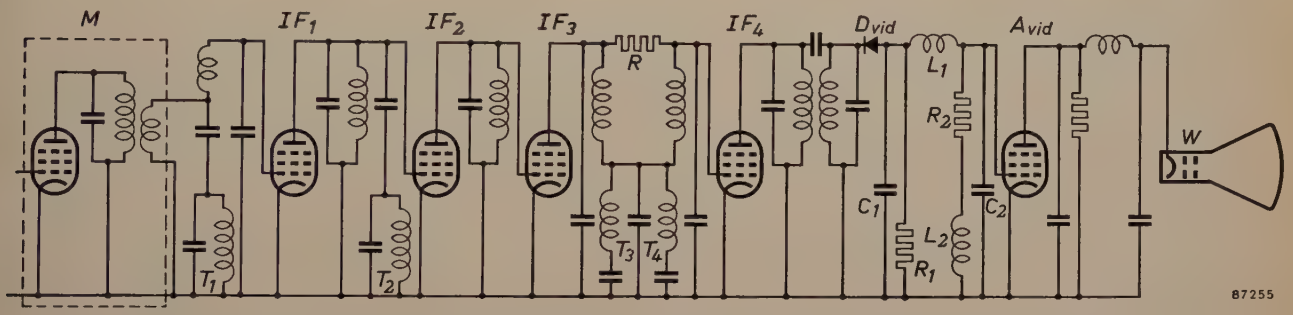


Fig. 17. Essentials of the circuit diagram of a phase-linear receiver. M mixer stage. IF_1-IF_4 IF stages. T_1 trap for suppressing the adjacent picture carrier (of frequency 31.9 Mc/s, P'' in fig. 4*b*). The effect on the trap circuit T_2 is to produce a sharp rise in the selectivity curve at 41.4 Mc/s (see curve I in fig. 18*a*, extreme right) and thus to improve the selectivity around that frequency. The trap circuit T_3 suppresses the sound carrier of the adjacent channel (40.4 Mc/s, Q') and the trap circuit T_4 suppresses the sound carrier of the channel being received (33.4 Mc/s, Q). The additional coupling provided by resistor R reduces the damping of T_3 and T_4.

The amplifier consists of five stages in two groups, a group of three stages (M, IF_1 and IF_2) and a group of two stages (IF_3 and IF_4). Each of the groups possesses a flat group delay curve within the pass-band and an amplification that falls by a factor of $\sqrt{2}$ at the limits of the band.

Apart from the trap circuits, each group contains four tuned circuits. The first group consists of a single stage with a band filter plus two further stages each containing one tuned circuit; these two are staggered. The two stages of the second group are each provided with a band filter.

Suppression of the adjacent and desired sound channels is effected by connecting two series resonant circuits (T_3 and T_4 in fig. 17), tuned to the frequencies concerned, in parallel with the coupling capacitor of the band filter of the last stage but one.

In order to confine the effect of T_3 to a narrow frequency band it is necessary to give high impedance to the components forming this filter. But by doing so the series resistance is also increased; to obtain effective suppression of the resonance frequency in spite of this, a bridging resistor R is employed to provide an additional coupling that is opposite in phase to the remaining coupling via T_3. In this way the effective Q of T_3 can be raised to as high a value as desired, at least insofar as the residual coupling at resonance is concerned.

The resistor R also effectively reduces the damping in the series-resonant circuit T_4, whose function is to suppress the sound carrier of the desired channel. The degree of suppression must not be excessive, on account of the employment here of the intercarrier sound system; the necessary amount of damping of T_4 is achieved by inserting an extra series resistance.

It will be seen from the circuit diagram that the first three stages are provided with the trap circuits T_1 and T_2. Their purpose is to give supplementary suppression in a certain range of frequency; T_1 does so at 31.9 Mc/s, i.e. at the frequency of the adjacent picture carrier, and T_2 increases selectivity beyond the adjacent sound carrier.

The resulting selectivity and group delay characteristics are shown in *figs. 18a* and *b* respectively as the broken-line curves marked I; for comparison, the corresponding curves (II) for a receiver of normal design are also given. These latter are the same as those in fig. 7. The difference between the properties of the two amplifiers is evident.

Even more interesting than the curves themselves is, of course, the quality of the picture received. Before discussing this question it is worth mentioning that the delay curves of the remaining receiver sections, i.e. the HF and VF sections, were also

covered in the investigation. The HF section proved, as was indeed to be expected, to have but little influence on the phase properties of the whole.

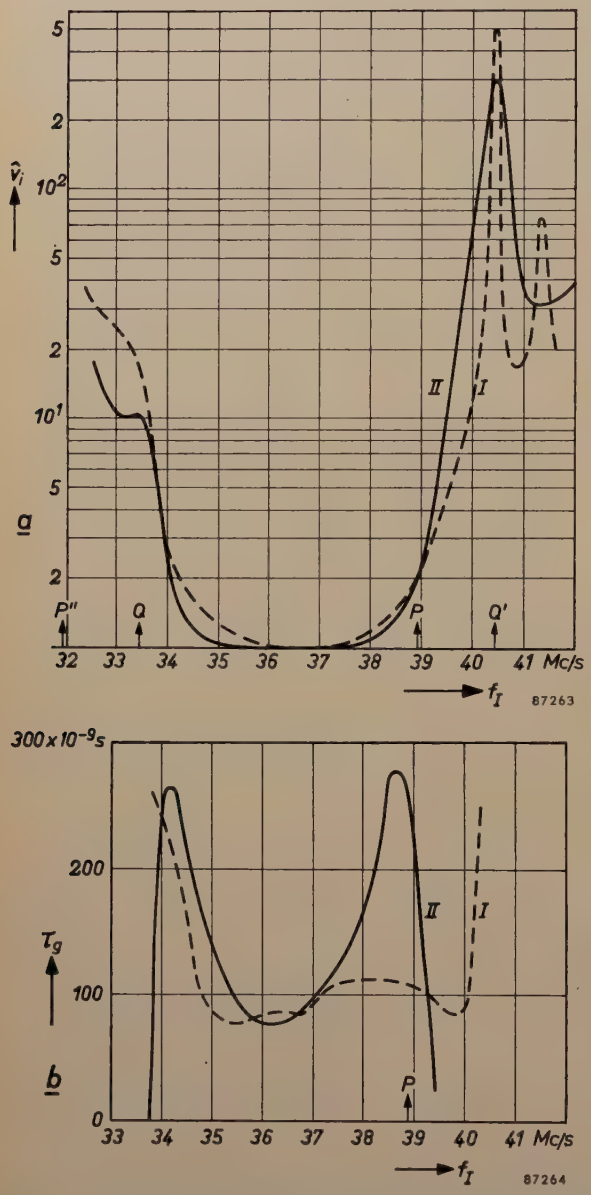


Fig. 18. *a*) Selectivity curve and (*b*) group delay characteristic, *I* for the phase-linear receiver and *II* for a normal receiver (curves *II* are taken from fig. 7).

There are few tuned circuits in this section of the receiver; no abrupt changes occur in the shape of the amplitude curve, and there are likewise no great delay variations.

The VF section contains, in the detector circuit, a network whose properties are not theoretically adequately understood [9]). However, the practical realisation of satisfactory amplitude and delay characteristics for this section offered no particular difficulties.

[9]) Philips Res. Rep. **10**, 291-292, 1955 (No. 4).

Practical results

Characteristic curves of a phase-linear and a normal receiver

Figs 19 and *20* show VF amplitude and VF phase delay characteristics measured on a phase-linear receiver (curves *I*) and a normal receiver (curves *II*). During the measurements the receivers were fed with a high frequency signal modulated with a VF voltage of variable frequency. Measurement were taken of the amplitude and group delay characteristics from the modulator to the control electrode of the picture tube; the phase-delay curve has been derived from the group-delay curve by graphical integration. The receivers were tuned exactly to the carrier frequency.

In both diagrams a further curve *III* is shown; it applies to the normal receiver subsequent to the incorporation in its VF section of the system of coarse phase-compensation mentioned on p. 40. It can be seen that this results in considerable dis-

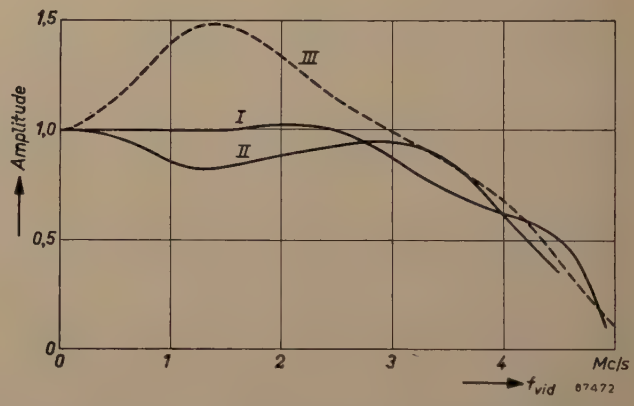


Fig. 19. Amplitude characteristics, *I* for the phase-linear receiver and *II* for a normal receiver. *III* amplitude characteristic for the normal receiver with rough phase compensation in its video frequency section.

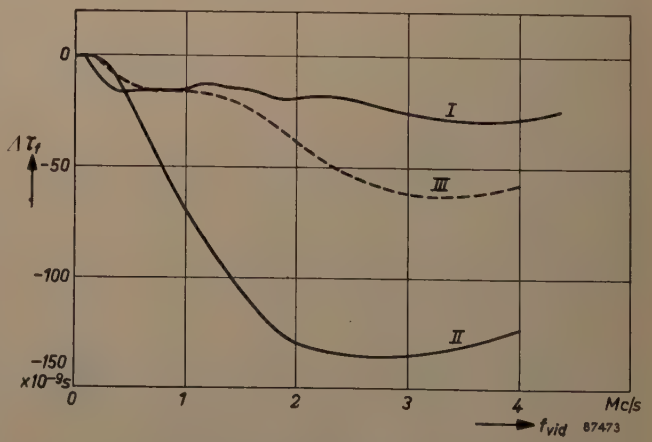


Fig. 20. Phase delay characteristics, *I* for the phase-linear receiver, *II* for a normal receiver and *III* for the normal receiver with rough phase-compensation in its video-frequency section. ($\Delta\tau_f$ is the difference with respect to the delay-time at low frequencies).

tortion of the amplitude curve while, with regard to phase-linearity, the result is not so good as that of the special receiver.

The characteristics shown in figs 19 and 20 were measured using an HF signal with a double sideband. In fact, the signal of a television transmitter is modulated with an asymmetrical sideband. The rough video frequency curves measured in the latter case differ slightly from those of figs 19 and 20. The difference is not appreciable until frequencies in the region of 1 Mc/s and is more noticeable for the phase-linear receiver than for the conventional receiver. The presence or absence of a second complete sideband has practically no effect on the conventional receiver by reason of the steep slope of the IF amplitude characteristic (see figs 18*a* and 4), while with a less steep slope the presence of signal components with frequencies near 40 Mc/s has some effect on the video signal after detection.

The rough amplitude curve of a phase-linear receiver measured when a vestigial sideband transmitter is used shows a narrow minimum of 10 to 15% at frequencies near 1 Mc/s. It can be shown that a distortion of this magnitude in the amplitude characteristic is hardly visible in the image [10]); this is also demonstrated by the fact that the pronounced distortion of curve *III* in fig. 19 is still acceptable. In any case this minimum can be compensated very easily in the video frequency section.

Step function response and the effect of modulation depth

Fig. 21 shows the response of the phase-linear receiver to step functions, as recorded by an oscillograph. The high-frequency signal was modulated with a square-wave voltage to a depth of 15% and 70%; fig. 21 shows the shape of the control voltage on the picture tube for the two cases.

Fig. 21. Step function responses of the phase-linear receiver (oscillograms of the control voltages on the picture tube) for white-to-black transitions (left) and black-to-white transitions (right), the modulation depths m being 15% in the upper diagram and 70% in the lower one.

The reason that the modulation depth has an effect is due to the use of an asymmetric side-band system for television. On detection, this produces distortion which is negligible when the modulation depth is small but becomes noticeable at large modulation depth. The kind of distortion produced in the case of a wave modulated with a square-wave voltage is illustrated in *fig. 22*; it may be seen from

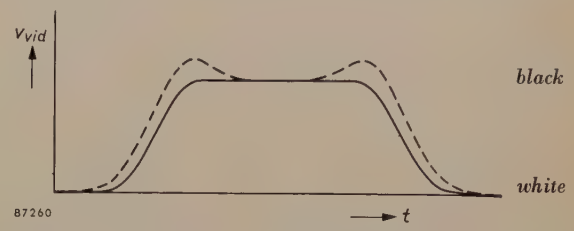


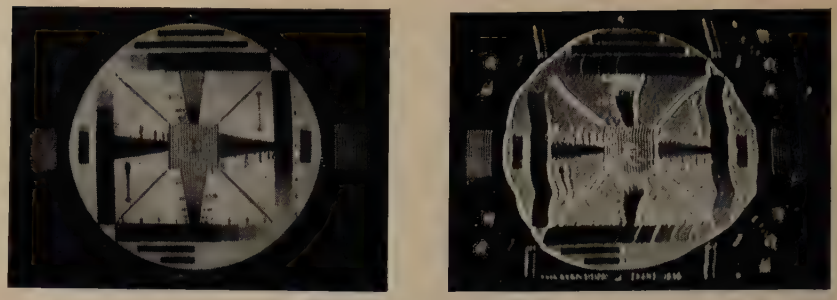
Fig. 22. As a consequence of the distortion arising from the use of an asymmetrical side-band system, the full curve of the step function response takes on the form of the dotted line. The distortion grows with increasing modulation depth and, in the phase-linear receiver, the effect it has can be useful (see fig. 21).

the figure that this effect in itself gives rise to smears with sharp transitions from *black to white* and thus aggravates the effect of phase distortion. The effect of single side-band distortion on a transition from *white to black* is precisely the opposite (fig. 9) and it more or less compensates for the phase distortion.

If the oscillograms for 15% modulation depth in fig. 21 are more closely examined, it will be seen that, apart from a slight amount of "ringing", they have a good measure of central symmetry; there is an equal amount of undershoot and overshoot. The central symmetry demonstrates the absence of phase distortion, for the latter deprives the step function of its property of central symmetry (cf. fig. 9). In the oscillogram for 70% modulation depth, it can be seen that the transition from white to black starts sharply; this demonstrates that the undershoot of the curve is compensated here by the distortion due to asymmetrical sideband rectification. On the other hand, at the black end of the transition the overshoot is reinforced; but this has no effect on the received picture, in which "blacker than black" cannot be seen. Something similar occurs for the transition from black to white; at the black end the undershoot is reinforced (which does not matter) and at the white end the overshoot is compensated as a result of single sideband distortion, and this greatly helps to improve the quality of the picture. All in all, the response of the receiver to step functions may be regarded as almost optimum.

[10]) A. van Weel, J. Brit. Instn. Rad. Engrs. **16**, 271-282, May 1956.

This happy state of affairs is the consequence of two favourable circumstances in combination: phase errors are obviated and, in connection with this, a "Nyquist flank" (the flank of the selectivity curve in the vicinity of the carrier frequency) of moderate steepness is employed (see fig. 18*a*, curve *I*). The steeper the Nyquist flank is, the more the asymmetrical side-band system assumes the properties of a pure single side-band system, with the distortion on detection that goes with it. Use of a Nyquist flank that is not so steep gives properties that tend towards those of a double side-band system, at least for modulation components of the lowest picture frequencies, and it is these that are most strongly represented in a step function and that give the greatest modulation depth. In such a case, therefore, asymmetrical side-band distortion is slighter; in this receiver, as we have seen, it is just enough to compensate overshoot and undershoot at the white end of the transition. It is of no importance that, at small modulation depths, undershoot and overshoot phenomena up to about 5% persist at both ends, for 5% represents a fraction of the reduced jump in brightness values that is no longer noticeable.



a Fig. 23. $\Delta f = +0.55$ Mc/s. *b*

a Fig. 24. $\Delta f = +0.50$ Mc/s. *b*

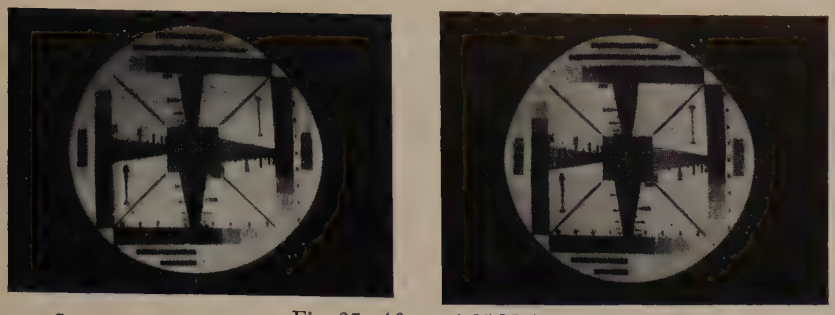
a Fig. 25. $\Delta f = +0.25$ Mc/s. *b*

a Fig. 26. $\Delta f = 0$. *b*

Figs 23-30. Photographs of screen images given by (*a*) the phase-linear receiver, and (*b*) a normal receiver having phase compensation in its video-frequency section. Δf: detuning of the transmitter (when Δf is negative, the picture carrier is shifted towards the centre of the pass-band). The generally poor definition of the images is due to their reduced size and to limitations in the printing process; with good tuning (as in fig. 26), it was possible to distinguish more than 400 lines on the screen with the naked eye.

Effect of detuning

Apart from optimum picture quality, the phase-linear receiver has a very useful property that conventional receivers only possess to a much slighter degree: detuning the receiver is found to have hardly any effect on the picture. With a receiver of the type used for the comparison in figs 18, 19 and 20, the picture quality is very sensitive to the tuning. It is not difficult to see why this should be bearing in mind that, when the conventional receiver is correctly tuned, a satisfactory picture is given only by virtue of the phase compensation in the video section; on detuning, and thereby changing the intermediate frequency of the carrier, the resulting IF phase errors will bring about quite different phase errors in the modulation (see fig. 18*b*), which are no longer balanced by the phase predistortion of the transmitter. This may be seen in *figs 23* to *30*. These are photographs of the picture screens of both receivers; the tuning of the receivers was left unchanged, the transmitter frequency being altered. The *a* series of photographs refers to the



a Fig. 27. $\Delta f = -0.25$ Mc/s. *b*

a Fig. 28. $\Delta f = -0.50$ Mc/s. *b*

a Fig. 29. $\Delta f = -1.25$ Mc/s. *b*

a Fig. 30. $\Delta f = -2.25$ Mc/s. *b*

(The photos of figs. 23-30 were taken during reception of a *double*-side-band transmitter. With reception of a single-side-band transmitter, the images given by positive detuning would be the same. For negative detuning — i.e. towards the middle of the pass-band — the effect would be different from that illustrated above, one reason being that the higher picture frequencies would be attenuated by half; this would naturally be the case for both receivers.)

phase-linear receiver and the *b* series to the ordinary receiver.

Phase-linear reception compared with reception using phase compensation

To conclude this article we shall once again weigh up the advantages and drawbacks of the two practical solutions to the phase problem.

We shall leave the question of phase compensation in the video section out of consideration, since the coarse compensation system described on page 40 cannot provide optimum picture quality, while ideal phase compensation having no effect on the amplitude characteristic cannot be employed in ordinary receivers for economic reasons. We shall confine ourselves, therefore, in making this comparison, to phase compensation in the video-frequency section of the transmitter, and the employment of phase-linear receivers.

One of the most important questions is whether an inherently phase-linear receiver is dearer to produce than a similar receiver of conventional type. This has proved *not* to be the case. The phase-linear receiver described above possesses the same number of stages and tuned circuits in its IF section as the receiver chosen for comparison, and the overall amplification given by its IF amplifier is likewise the same. The alignment of the IF circuits in the phase-linear receiver will be less critical than that in a normal receiver, in which large errors are compensated by intentionally introducing opposite errors, because, in the latter system, small relative variations can produce big absolute errors.

Further striking advantages offered by the phase-linear receiver, as against the system of compensation at the transmitting end, are as follows.

1) Picture quality is much less sensitive to variations in tuning.

2) Optimum picture quality is obtained as a consequence of the reduced effect of vestigial side-band detection.

3) The unfavourable influence of a parasitic phase modulation at the transmitter is less noticeable the flatter the I.F. amplitude curve at frequencies near the carrier frequency. For this reason the phase-linear receiver is less

sensitive to such parasitic phase modulations.
4) No artificial phase distortion needs to be introduced into simple receivers without traps, intended for local reception, in order to compensate predistortion at the transmitter, which is pointless in such circumstances.
5) If phase pre-distortion can be dispensed with at the transmitter, the modulation is less difficult: the pre-distortion (or "pre-emphasis") means sending out "whiter than white" to compensate for the black smears in the received picture, and therefore it necessitates a greater maximum modulation depth in the transmitter. This increased modulation depth increases the distortion which accompanies detection of an asymmetrical side-band signal.
6) Both in the production of television receivers and in servicing workshops, picture quality is usually checked with the help of a standard signal generator, the signal from which is double side-band modulated with a testing signal. If the receiver's phase errors are compensated at the transmitter, corresponding phase predistortion will have to be incorporated in the testing generator. This complication does not affect the production of phase-linear receivers, and a much simpler generator can therefore be used [10]).

One disadvantage of the phase-linear receiver discussed here is that the suppressed frequency band is narrow (fig. 18*a*). Consequently, if interference from an adjacent transmitter is present, fairly high demands have to be made as to the stability of the local oscillator. That is to say, the frequency of the interfering signal in the IF amplifier must never be allowed to deviate from that to which the trap circuit is tuned.

Hence the aim of continuing development work is to widen the suppressed band. It may be mentioned that, at the time of writing, success has been obtained in extending the suppressed band to about 300 kc/s, with attenuation to 1%. Such a width allows for acceptable tolerances as regards the stability of the local oscillator. The circuit necessary for this increased width of the suppressed band is not complicated and does not involve higher production costs.

Another objection that might be made to the phase-linear receiver described here concerns the falling off of the amplitude characteristic at high modulation frequencies. As mentioned above, a compromise must be sought here between greater steepness of the step function response and a higher degree of overshoot. We settled for a falling off of the amplitude characteristic by one half at the edge of the pass-band. However, there are receiver designers who want the amplitude curve to run flat right up to the cut-off frequency and who accept the correspondingly high overshoot factor (9%) for the sake of obtaining a flank of somewhat steeper slope. It is quite feasible to put this principle into practice; but then, of course, the group delay curve at high modulation frequencies will start to climb earlier and more rapidly. This however has little effect on the phase delay of the modulation, as already stated. The delay curve shown in fig. 8*b* was measured on an IF amplifier that possesses a flat amplitude characteristic up to 34.4 Mc/s; thus its video-frequency amplitude characteristic is flat up to 4.5 Mc/s. As may be seen from fig. 8*b*, considerable deviations in group delay have but little effect on the phase delay. An amplitude characteristic of this kind can therefore be realized fairly simply. Difficulties as to phase-linearity are mainly to be expected in the vicinity of the carrier frequency.

In general, the obviation of errors is to be preferred to their compensation by the artificial introduction of opposite ones, unless this offers definite advantages. In the present instance we know of no advantages of this latter practice, but there are several disadvantages as have been outlined above. We believe therefore that the transmitter should radiate a signal that contains as little amplitude and phase distortion as possible, and receivers should likewise be designed such that a minimum of amplitude and phase distortion arises in them. This point of view seems to us to be not only logical but also quite capable of being put into practice; moreover, taking the long view, it does not tie one down to the present-day stage of development. An additional advantage is that it does away with the need to supplement the Gerber Standard with an internationally laid down system of phase compensation. There is no mention of phase properties in the Gerber Standard as at present formulated. This may be interpreted, with regard to the future, as a stipulation that both transmitter and receiver should be phase-linear, in the same way that both should exhibit a flat amplitude characteristic.

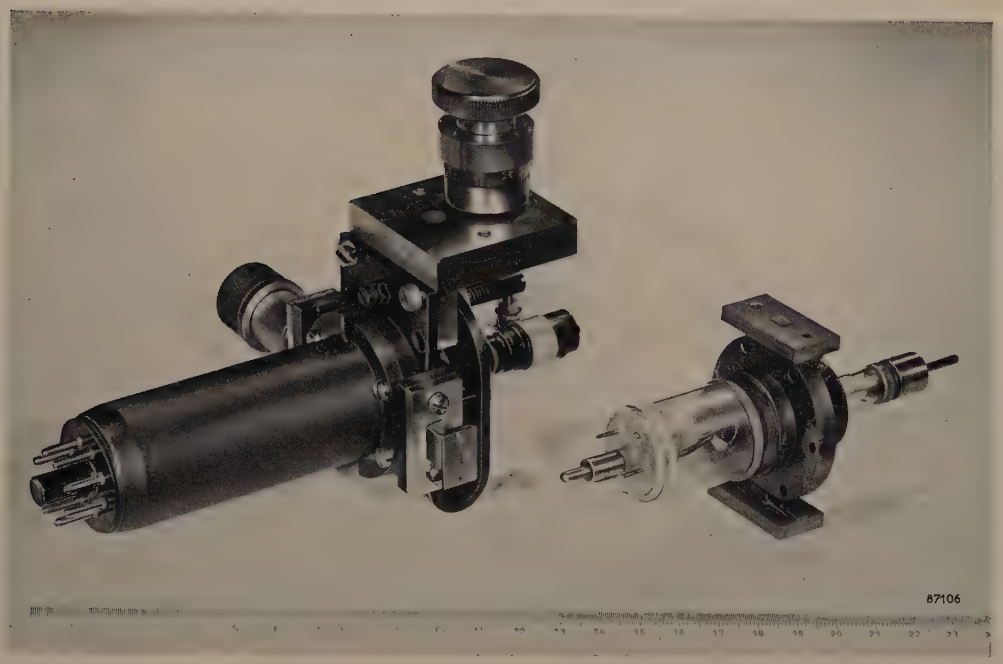
Summary. The growing number of television transmitters has made it necessary to increase the selectivity of receivers especially in the IF section. This has entailed considerable deviations from the linear relationship between phase and frequency necessary to provide a picture of good quality. Up to now a solution of this problem has been sought in the compensation of phase errors, either in the video section of the receiver (where, to give optimum results, it would be very expensive) or in the video section of the transmitter. Phase compensation at the transmitter answers the purpose to only a limited extent,

for the various types of receiver differ widely in their phase properties.

Instead of compensation by the introduction of opposite errors, the writer argues for a system in which phase errors can be largely avoided. An IF amplifier was aimed at which would not only possess an optimum amplitude characteristic within the pass-band and satisfy demands as to selectivity outside it, but also be very nearly phase-linear within the band. The aim was to realize such an amplifier without an increase in the cost of manufacture. This aim has been achieved, making use of Bode's theorem with a 5-stage amplifier, the amplitude curve of which falls to 50% of maximum amplification at the limits of the passband; the requisite selectivity at frequencies capable of causing interference is provided by four trap circuits. In this way optimum picture quality has been obtained for the given bandwidth. An additional favourable property of the phase-linear receiver lies in the fact that its picture quality is much less sensitive to detuning than is the case in an ordinary receiver; this is demonstrated by photographs of screen images.

A REFLEX KLYSTRON FOR 4-mm WAVES *)

621.373.423.029.65



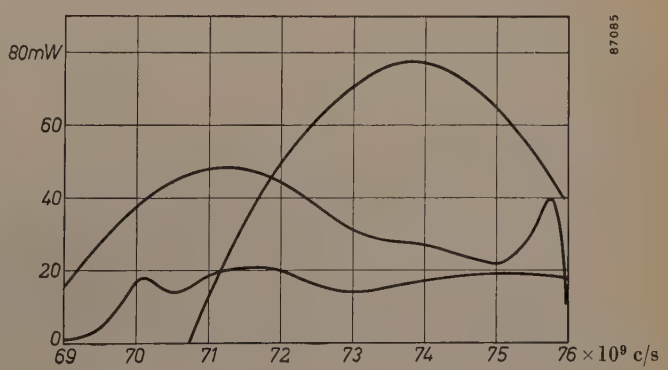
In the wake of centrimetre waves, which in the past twenty years have been widely applied, particularly in the field of radar, it is now the range of millimetre waves that is commanding increasing interest. This interest is partly attributable to the wish to improve the "resolving power" of short-range radar installations, partly also to the desire to explore this field with a view to new possible applications.

The shortest wavelength employed in practice so far is 8 mm. This restriction is mainly imposed by the limitations of the available oscillator tubes.

A considerable step forward in the attempt to open up the range of millimetric waves for practical application has been made at the Eindhoven Research Laboratories by the development last year of a reflex klystron for 4-mm waves, giving a continuous output of some tens of milliwatts.

The photograph (scale in centimetres) shows on the right the klystron itself and on the left an assembly fitted with a tuning device, matching plunger, output circuit and base.

The principal difficulties in the construction of tubes for these extremely short waves are due to the necessarily very small size of all component parts.



*) This note first appeared in the Dutch edition of this review: Philips Technisch Tijdschrift **18**, 55-56, February 1956.

The separation between reflector and resonator gap, for example, is only 0.2 mm in the 4-mm tube, across which a voltage of 2.5-3 kV has to be applied. The current-density required of the cathode, which has an emitting surface only about 1 mm in diameter, is roughly 2 A/cm². The construction of such a tube, with a reasonable working life, was made feasible only by the use of the L-cathode. A more comprehensive description of the tube will shortly appear in this Review.

The graph shows the output power plotted as a function of the frequency for three specimens of the tube. The disparity in the results is admittedly still fairly large, but the power outputs attained at bandwidths of 7-10%, are ample for a number of applications.

B. B. van IPEREN.

THE STUDY OF THERMAL CONDUCTIVITY PROBLEMS BY MEANS OF THE ELECTROLYTIC TANK

by F. REINIGER *).

536.212:621.386.1.032.22:621.317.329

The electrolytic tank, a well-known aid in the study of electrostatic fields, may also be successfully applied for the study of steady state thermal conductivity problems. This article describes an investigation, using this technique, into the thermal behaviour of an oil-cooled copper X-ray anode with a small thin plate (pastille) of tungsten embedded in it.

It is well-known that the form of electric equipotential surfaces between an arbitrary electrode system can be determined by means of the electrolytic tank [1) 2) 3)]. This has shown itself to be of great practical use for electrode systems which are of such an involved nature that direct calculation of the field is impracticable. *Fig. 1* shows a possible arrangement in diagrammatic form. A sinusoidal alternating voltage originating from the oscillator *I* is applied to the electrodes E_0 and E_1 of the electrolytic tank and a potentiometer P_0P_1 connected in parallel. The probe electrode *S* is dipped in the electrolyte and the position of the sliding contact on the potentiometer is adjusted so that the amplitude of the alternating voltage on this contact is equal to that on the probe. Keeping the sliding contact fixed, the probe is then moved along the surface of the liquid so that its potential remains the same. The probe then describes an equipotential line which can be recorded on a drawing board set up near the tank by means of a stylus coupled mechanically to the probe. The method best lends itself to the study of two-dimensional problems but it can quite simply be extended to cover three-dimensional ones with a plane of symmetry, the surface of the liquid in this case having to coincide with the plane of symmetry.

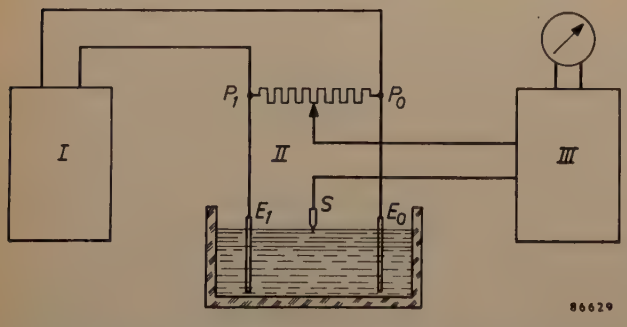


Fig. 1. Diagram of an arrangement for measuring the potential in an electrolytic tank. *I* oscillator, *II* tank and potentiometer, *III* amplifier with meter.

The study of electrostatic fields by means of the tank rests on the analogy between the dielectric displacement (vector **D**) in an electrostatic field and the current density (vector **j**) in a conducting medium. For the component of **D** and **j** in an arbitrary direction (l) we have the equations:

$$D_l = \varepsilon E_l = -\varepsilon \, \partial u/\partial l \quad . \quad . \quad . \quad . \quad . \quad (1)$$

and

$$j_l = \gamma E_l = -\gamma \, \partial u/\partial l \quad . \quad . \quad . \quad . \quad . \quad (2)$$

*) C. H. F. Müller Aktiengesellschaft, Hamburg. The investigation described in this article partly refers to experiments carried out in 1945 but about which nothing has so far been published.

1) G. Hepp, Potential measurements by means of the electrolytic tank, Philips techn. Rev. **4**, 223-230, 1939.

2) F. Möller, Arch. techn. Messen, V24-1, July 1947.

3) N. Warmoltz, Potential distribution of the igniter of a relay valve with mercury cathode, Philips tech. Rev. **8**, 346-352, 1946.

Table. Corresponding quantities in the electric field, the thermal field and the electrostatic field

Electrical		Thermal		Electrostatic	
Current I	[A]	Heat flow N	[W]	Charge Q	[As]
Current density j	[A/m²]	Thermal current density q	[W/m²]	Displacement D	[As/m²]
Potential U	[V]	Temperature T	[°C]	Potential U	[V]
Field strength E	[V/m]	Temp. gradient τ	[°C/m]	Field strength E	[V/m]
$U - U_0 = u$	[V]	$T - T_0 = \vartheta$	[°C]	$U - U_0 = u$	[V]
Conductivity γ	[(A/m²)/(V/m)]	Conductivity λ	[(W/m²)/(°C/m)]	Dielectric const. ε	[(As/m²)/(V/m)]
Specific. res. ϱ	[(V/m)/(A/m²)]	—		—	
Conductance G	[A/V = Ω⁻¹]	Conductance L	[W/°C]	Capacitance C	[As/V]
Resistance R	[V/A = Ω]	Resistance R_w	[°C/W]	—	
—		Heat transfer coefft. α	[W/m²°C]	—	

respectively, where $\varepsilon = \varepsilon_0 \varepsilon_r$ is the dielectric constant of the medium, **E** the field strength, $u = U - U_0$ the voltage with respect to a point at potential U_0 and γ the conductivity of the medium.

It is perhaps less well known that it is also possible to investigate steady state problems of heat flow by means of the electrolytic tank [4]) on the basis of the analogy between the expression (2) and that for the flow of heat **q**. Considering again the component in the direction l,

$$q_l = \lambda \tau_l = -\lambda \partial\vartheta/\partial l \quad . \quad . \quad . \quad . \quad (3)$$

In this expression λ is the coefficient of thermal conductivity of the medium, $\vartheta = T - T_0$ the difference between the local temperature T and a chosen, fixed temperature T_0 and τ the temperature gradient $= -\partial\vartheta/\partial l$ (in the direction of the most rapid decrease).

To clarify the analogy and the subsequent considerations of the thermal problem investigated, corresponding quantities (and their units) are given in the Table for the electric field and for the thermal field; for the sake of completeness, the corresponding quantities in the electrostatic field are also shown. The correspondence follows from a comparison of the differential equations for the three cases, as given in the appendix to this article.

The analogy was used in the present instance for investigating the thermal behaviour of the anode of an X-ray therapy tube (type 71771/05, 200 kV, 4 kW continuous load; see fig. 8). The anode consists of a block of copper with a small, thin plate of tungsten on the front of it where the energy absorption takes place. The other side of the copper is cooled with oil and the heat thus conducted away.

The object of this experiment was to provide details of the temperature distribution in the anode, in relation to the thickness of the copper block, and more particularly, of the distribution of the temperature over the surface in contact with the cooling liquid. Since tungsten is a poorer conductor of heat than copper, the dimensions of the tungsten plate affect the temperature distribution. The effect of the size of the focus on the temperature at the focus was also investigated.

Design of the model

In its simplest form, the electrolytic tank comprises a trough filled with a liquid in which two electrodes maintained at constant potential $U = U_0$, $U = U_1$ are dipped. Since $u = U - U_0$ is analogous to Q, the method can be used wherever heat is supplied or absorbed by constant temperature surfaces.

If the heat-conducting medium in the original consists of various materials with different values of thermal conductivity (see Appendix, equations (12) - (14), (16) - (17) and (21) - (23), the model must have areas corresponding to these materials having different electrical conductivities i.e. different electrolytes. The potential must be the same on both sides of the boundary between two areas, i.e. adjacent electrolytes must be separated from each other by a wall whose electrical conductivity is large perpendicular to the plane of the wall but whose conductivity is zero along the wall. In two-dimensional cases, use is made of a wall of insulating material provided with metal strips [3]) [5]) (*fig. 2*). At the position of the strips the potential on both sides of the wall is the same, while there is no conductivity along the wall since the strips are

[4]) J. Malavard and J. Miroux, Electrical analogies for Heat Transfer Problems, Proc. 4th Int. Congr. Industr. Heating, Paris 1952; see Eng. Digest **13**, 417-420, 1952.

[5]) R. Stachowiak, Störungsfreie Trennwände für zusammengesetzte Elektrolytmodelle. E.T.Z. **62**, 441-443, 1941.

insulated from each other. In other cases a wall is used which is pierced by a large number of rivets whose heads make contact with the liquids on both sides (*fig. 3*).

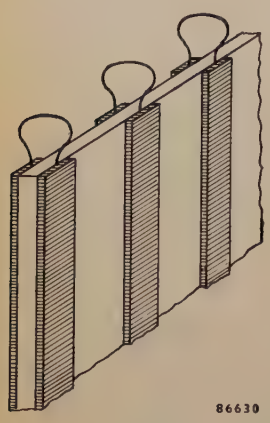


Fig. 2

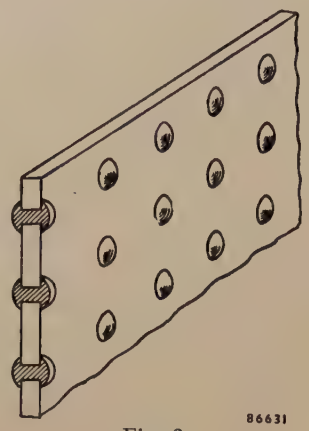


Fig. 3

Fig. 2. Construction of a dividing wall with local electrical contact between two electrolytic conductive liquids (two-dimensional).
Fig. 3. The same as fig. 2 but for a three-dimensional case.

In our particular case another method is used to produce media of different conductivities. Use was made of the principle that a sponge of non-conductive material soaked in an electrolyte is less conductive than the electrolyte liquid itself. The decrease in conductivity resulting from the increase in the current path and the decrease in the conducting cross-section between two points can be measured using the apparatus shown in *fig. 4* where the resistance of the electrolyte is determined firstly without, and then with the sponge.

The resistance can be easily increased by a factor of up to 8 by a judicious choice of porosity. There are a number of porous substances suitable for this purpose, some elastic (like rubber) and others non-elastic but all having widely differing porosities (pore volume/total volume). It is of course essential that all pores are interconnected and are distributed isotropically.

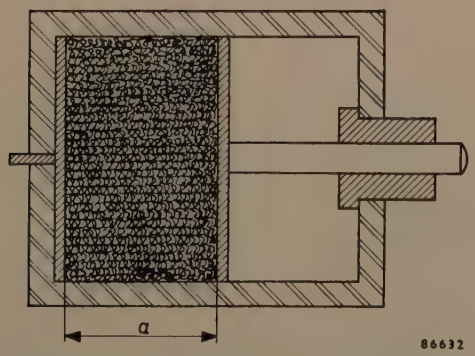


Fig. 4. Device for determining the resistance of an electrolyte and of a sponge-like mass saturated with electrolyte.

It was also possible in the apparatus given in fig. 4 to investigate the manner in which the resistance varies as the sponge is compressed thereby reducing the porosity. The result of this measurement is reproduced in *fig. 5*. From these results, it can be seen that up to 50 × the original specific resistance can be obtained by compression but care must be taken that the compression does not spoil the isotropy of the sponge. Since it is not easy to work with compressed sponges in a model, this method was not used in the present tests.

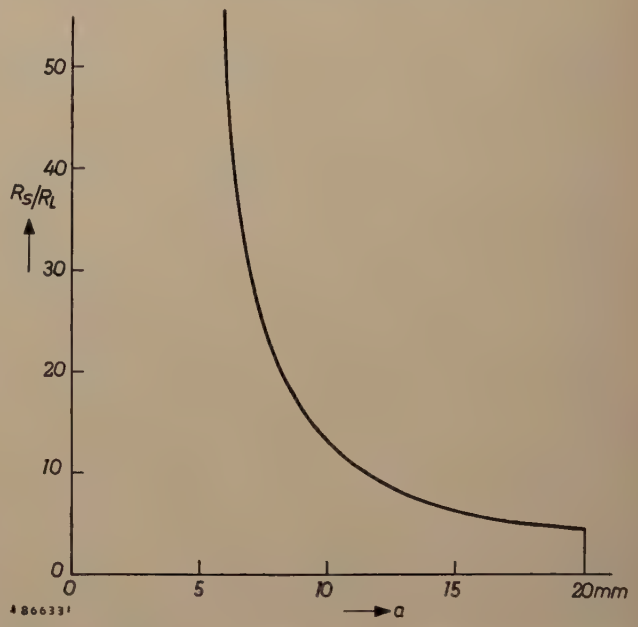


Fig. 5. Variation of the resistance of a sponge saturated with electrolyte with compression. The increase in resistance with respect to the resistance of the electrolyte, R_S/R_L, is plotted as a function of the length a in figure 4. In the non-compressed state (a = 20 mm), $R_S/R_L \approx 4$.

It is important that the pores, should be completely filled with liquid. Any air-bubbles left behind would give additional resistance to the current. With compressible masses this can be achieved by immersing the sponge completely in the electrolyte and compressing it repeatedly. Non-elastic porous materials can be saturated with liquid in a vacuum and have the advantage of making the model somewhat more rigid.

A condition demanding special attention in the analogy between a thermal and an electric field is the occurrence of temperature discontinuities. In practice, for example, discontinuities occur when a metal wall is being cooled by means of a cooling liquid. There is a layer of liquid which is a bad conductor between the wall to be cooled and the cooling liquid which leads to a temperature drop $\vartheta_1 - \vartheta_2$, such that the local heat flow is given by

$$q = \alpha(\vartheta_1 - \vartheta_2) \ldots\ldots \quad (4)$$

The quantity α is called the heat transfer coefficient It is customary to take the temperature of the inflowing cooling liquid for ϑ_2. The heat transfer inside the cooling liquid is then included in the constant α.

A layer of this kind can be simulated in the model by a perforated non-conducting plate (*fig. 6*) which can be considered as an anisotropic sponge. If the plate contains n holes per unit area and these holes each occupy an area of $A = \pi r^2$ (r = radius of hole) then, if δ is the thickness, the current intensity becomes:

$$j = (n\gamma A/\delta)(u_1 - u_2).$$

The quantity α corresponds to $n\gamma A/\delta$ and consequently we must ensure that: $\alpha/\lambda = (n\gamma A/\delta)/\gamma = nA/\delta$.

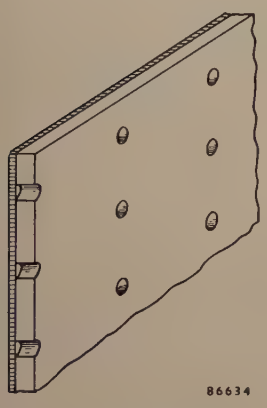


Fig. 6

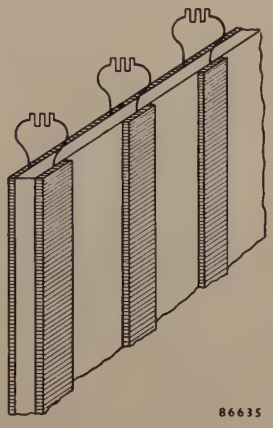


Fig. 7

Fig. 6. Method for obtaining a potential step between the liquid and an electrode by means of a perforated insulating plate covering the electrode (three dimensional). The potential step is the electrical analogy of the temperature drop due to heat transfer across an interface.
Fig. 7. Method for obtaining a potential step between the liquid and an electrode by means of a number of resistances. The arrangement shown is for a two-dimensional case.

Another method of achieving the same thing is to provide the plate with a large number of small metal strips or buttons mounted on an insulating wall (in a two-dimensional case we can again resort to strips, see *fig. 7*). each connected via external resistors to an electrode on the opposite side of the wall. Although this arrangement is rather complicated with large models, since the resistors must have values adapted to whatever electrolyte is being used, it has the advantage of enabling the value of α to be adjusted as required and also of allowing it to have a different value at a particular place. (The same can be accomplished with the drilled plate by varying the diameter of the holes.)

As α depends upon the kind of cooling liquid and upon the state of flow, which is also determined by the geometry of the cooling space, the value of α must be ascertained in advance for complicated cooling systems by separate experiments under conditions which simulate as closely as possible those in the original.

The way in which the model is used depends upon the known data e.g. the boundary conditions. If in the original the temperatures at the boundaries be given, then the potentials are adjusted accordingly. For example, let there be two temperatures T_1 and T_0 ($T_1 > T_0$) and let U_1 and U_0 be the corresponding potentials. By ascertaining the equipotential line (in the plane in which the probe moves) information is obtained enabling the temperature at each point to be known. We have:

$$\frac{T - T_0}{T_1 - T_0} = \frac{U - U_0}{U_1 - U_0} \quad . \quad . \quad . \quad . \quad . \quad (5)$$

or

$$\frac{\vartheta}{\vartheta_1} = \frac{u}{u_1} . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

In other cases (such as the above-mentioned X-ray anode, to which we shall presently refer again in more detail) the total heat flow is known and what is required is the difference in temperature between two points (viz. the focus and the cooling liquid). For this case, it is useful to introduce the conception of thermal resistance (temperature difference/heat flow). In the model, an electrical resistance corresponds to this. Since at a particular point x, y, z in the original, we have for the x component of the heat flow:

$$\Delta N_x = q_x \Delta y \Delta z = -\lambda \Delta y \Delta z \cdot \Delta\vartheta/\Delta x, \quad . \quad . \quad (7)$$

and at the corresponding point x', y', z' in the model:

$$\Delta I_{x'} = j_{x'} \Delta y' \Delta z' = -\gamma \Delta y' \Delta z' \cdot \Delta u/\Delta x', \quad . \quad (8)$$

we obtain

$$\frac{\Delta\vartheta}{\Delta N_x} = -\frac{\Delta x}{\lambda \Delta y \Delta z} \quad \text{and} \quad \frac{\Delta u}{\Delta I_{x'}} = -\frac{\Delta x'}{\gamma \Delta y' \Delta z'}$$

and therefore

$$\frac{R_w}{R} = \frac{\text{thermal resistance}}{\text{electrical resistance}} = \frac{\gamma}{\lambda}\frac{\Delta x}{\Delta x'} \cdot \frac{\Delta y' \Delta z'}{\Delta y \Delta z} = m\frac{\gamma}{\lambda},$$

where m represents the scale ratio:

$$m = \frac{\Delta x'}{\Delta x} = \frac{\Delta y'}{\Delta y} = \frac{\Delta z'}{\Delta z}.$$

Application to the X-ray Anode

We will now show how the preceding considerations are applied in practice in a particular case. The anode of the X-ray tube mentioned above which

is illustrated in *fig. 8* was the object to be investigated. The electrons strike the focus in the centre of the small tungsten plate *W* [6]) and the heat produced is conducted via the tungsten and the copper block Cu to be finally absorbed by the cooling liquid *K*.

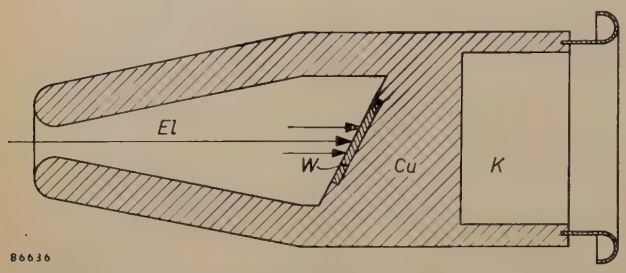


Fig. 8. Cross-section of the anode of an X-ray therapy tube with an oil-cooled anode. Anode potential 200 kV; continuous power rating 4 kW. *El* electron stream, *W* tungsten plate, *Cu* copper body of anode, *K* = cooling liquid flowing in turbulent motion.

The object of the investigation is to ascertain the effect on the temperature distribution of the design: in particular the effect of the thickness *A* of the anode (distance from centre of focus to cooling surface) and of the temperature at the focus with a given power supply.

On account of the symmetry with respect to the plane of the drawing in fig. 8, it is sufficient to consider only half the anode in the model (see *fig. 9a, b*). This is bounded on the outside by a cylindrical surface and since we can disregard the slight loss of heat due to radiation, it is justifiable to represent this surface in the model by an insulating wall. For the latter a sheet of transparent plastic material ("Perspex") bent round to form a half cylinder was used. This half cylinder was enclosed at both ends by a wall of insulating material so that a semi-cylindrical tank was produced and this was filled with electrolyte. The milling on the anode at the side of the tungsten plate was represented by an object (a truncated half cylinder) made of insulating material ("Cellon"). On its oblique face on the cooled side, was fixed a sand-blasted sheet-iron electrode which was to take the place of the focus (the purpose of sand-blasting is to clean the surface and improve the metal-liquid contact).

The model is based on the assumption that the temperature of the focus is uniform at every point but this is not the case in actual fact. The temperature distribution in the focus depends upon the distribution of the density of the electron stream. At constant current density, the temperature at the centre of the focus will be higher than at the boundary. This detail is not taken into account in the model. It can be said that the potential of the electrode corresponds to the mean focus temperature.

The outer surface of the cooling liquid was represented by a semi-cylindrical electrode, likewise of sand-blasted sheet-iron. On the side where the liquid was, this electrode was completely covered with a perforated, insulating sheet ("Cellon") which represents the heat transfer coefficient. The value of nA/δ used in the model was $a = 0.8$ W/cm² °C

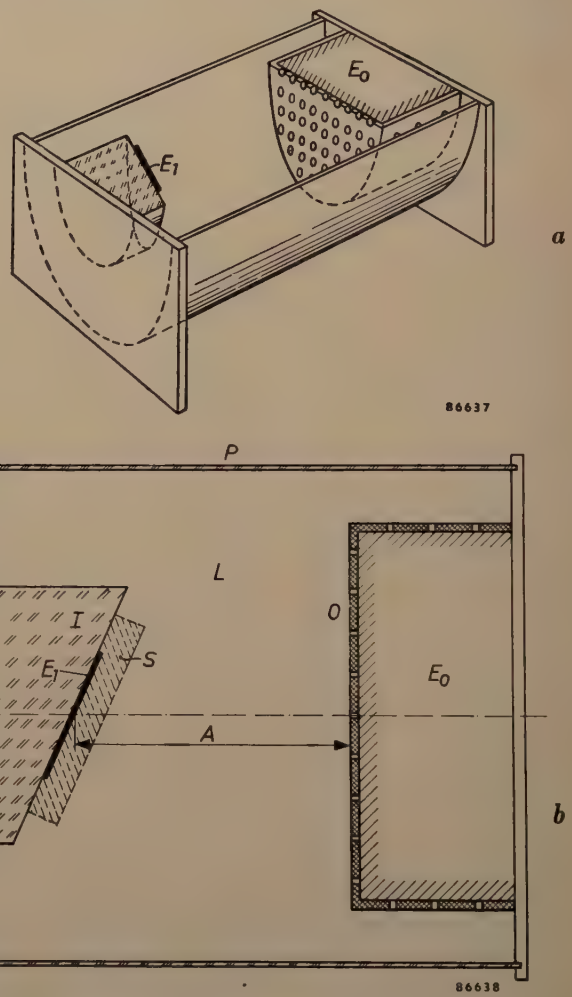


Fig. 9. Perspective sketch (*a*) and cross-section (*b*) of the electrolytic tank model of the X-ray tube anode. *L* electrolyte which simulates the heat conducting copper anode. *P* semi-cylindrical insulating outside wall of the anode, *I* insulator ("Cellon") representing the vacuum in which the electrons move towards the focus, E_0 and E_1 electrodes simulating the cooling liquid and the X-ray focus. *O* perforated insulating layer on E_0 representing the heat transfer coefficient *a*, *S* = sponge representing the tungsten plate. Three models were prepared, each having a different distance of *A*. (In actual fact the iron electrode E_0 and the insulator *I* did not need to be made full size but only as thin, conducting and non-conducting limiting walls respectively for the liquid *L*).

[6]) Tungsten is chosen on account of its large X-ray efficiency (high atomic number) and because of its high melting point and low vapour pressure at high temperatures. The relatively poor thermal conductivity of tungsten must be accepted and for this reason the tungsten is in the form of a small, thin plate (pastille) embedded in a substance of good thermal conductivity which may have a lower melting point (electrolytic copper).

(determined from previous experiments concerning the cooling of copper surfaces with oil flowing with turbulent motion).

In the first instance a 3 : 1 scale model was used (scale ratio $m = 3$), the model being in three designs corresponding to anode thickness A of 35, 25 and 15 mm. Since we are primarily concerned with determining effect of the anode thickness on the temperature distribution at the cooled side, nothing took the place of the tungsten plate in this model. Consequently the model corresponds to an anode consisting solely of copper. The measured equipotential lines for all three cases are drawn in *fig. 10*. The potential with respect to the cooling liquid is given in percentages of the total potential difference.

thermal conductivity of copper is 370 (W/m²)/(°C/m) We now find

$$R_w = m\,R\,.\,\frac{\gamma}{\lambda} = 3 \times 378 \times \frac{0.0445}{370}\,\frac{°C}{W} = 0.137\,\frac{°C}{W}$$

and therefore the thermal conductivity $L = 1/R_w = 7.3$ W/°C.

It follows from this that at a load of 4 kW the temperature difference between cooling liquid and focus is:

$$\vartheta_1 = 4000 \times 0.137\ °C = 550\ °C.$$

At a cooling liquid temperature of 50 °C, temperature T_1 of the focus is accordingly in the region of 600 °C.

The temperature of the copper of the cooling surface can now be easily derived from the per-

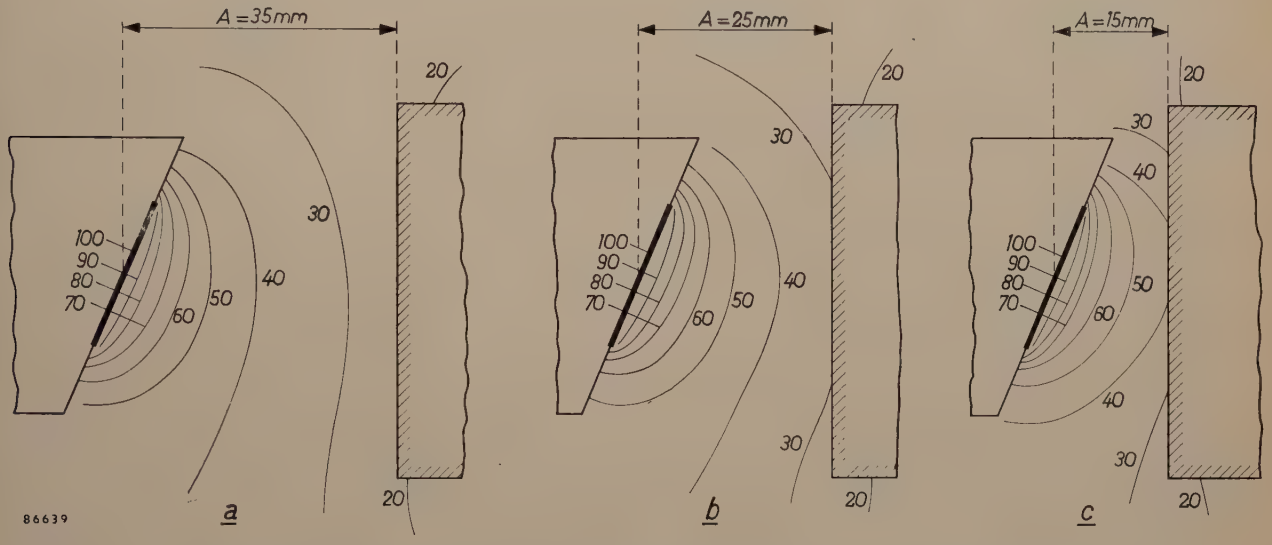


Fig. 10. Equipotential lines measured in a 3 :1 scale model of the copper anode without the tungsten plate for three thicknesses of the anode: *a*) $A = 35$ mm, *b*) $A = 25$ mm, *c*) $A = 15$ mm.

As can be seen, the rear wall of the anode when 35 mm thick is practically an equipotential surface. In the other cases the temperature varies fairly strongly along the rear wall so that this wall of the anode may in places be hotter than is permissible if decomposition (carbonisation) of the cooling liquid (oil) in the neighbourhood of the copper surface is to be avoided.

The thermal resistance of the anode can be derived from the model tests as follows.

With a thickness of $A = 35$ mm the electrical resistance R of the "half" model amounted to 757 ohms. For a complete model it would accordingly be half of this, or 378 ohms. The specific resistance $\varrho = 1/\gamma$ of the electrolyte was measured in an arrangement such as is shown in fig. 4. This came to 22.5 Ωm and therefore $\gamma = 0.0445\ \Omega^{-1}m^{-1}$. The centage values given in fig. 10. When $A = 35$ mm, for example, the maximum value of the potential near the cooling surface is about 0.25, so that the temperature of the copper comes to 50 °C + 0.25 × 550 °C ≈ 185 °C (temperature discontinuity approximately 135 °C).

Later work was done with a 10 : 1 scale model ($m = 10$) and in this model the place of the small tungsten plate was taken by a flat rubber sponge whose pores were saturated with electrolyte. The ratio between the coefficients of thermal conductivity of tungsten and copper at the temperatures which can be anticipated amounts to roughly 3.7:

$$\lambda_{Cu} = 370\ (W/M^2)/(°C/m),$$

$$\lambda_W = 100\ (W/m^2)/(°C/m).$$

Care was taken to ensure that the conductivity of

the electrolyte liquid was decreased in the same ratio (approx 1/4) by means of the sponge. In this model, various shapes (rectangular and circular) were used for the focus, i.e. for the electrode E_1.

In *fig. 11*, the equipotential lines are drawn for an anode 25 mm thick both with and without a tungsten plate. It was possible to obtain the temperature distribution along the normal to the focus from these measurements and the result is reproduced in *fig. 12*. It can be seen from this figure that the temperature curve in the copper with a tungsten plate is practically identical to the one without a plate and furthermore, we can see that the plate makes the temperature of the focus increase strongly. The temperature step (in this instance about 175 °C) between the rear wall of the anode and the cooling liquid can also be clearly seen.

Finally, *fig. 13* gives a résumé of the measurements on various models, with and without a tungsten plate and with different thicknesses of the anode. In these figures, the thermal conductivity

$$L = \frac{\text{power supplied}}{\text{difference in temperature between focus and cooling liquid}}$$

is plotted as a function of the thickness A (in mm) for various dimensions of the focus (in mm). Obviously, the conductivity is considerably less with the tungsten plate than without it (about half as much) and the conductivity is greater according as the power supplied is distributed over a greater area (larger focus).

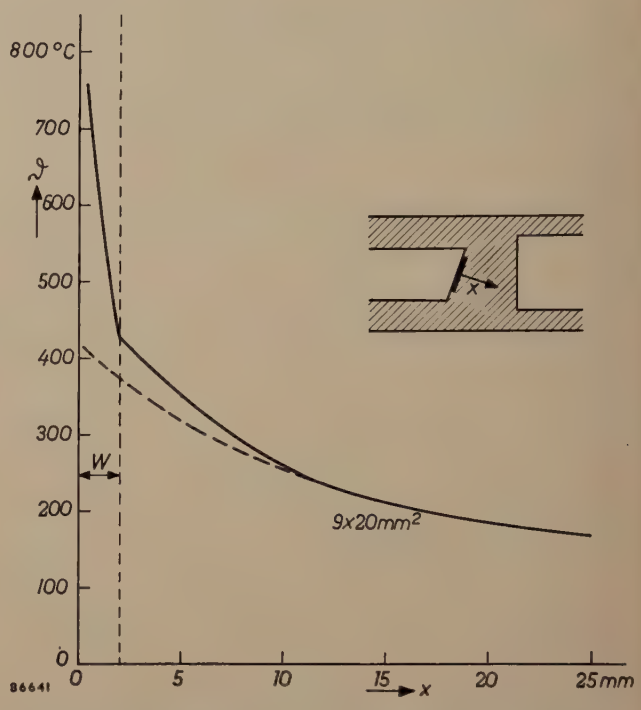


Fig. 12. The temperature curve along the normal in the centre of the focus, obtained from the measurements in fig. 11. The temperature difference ϑ with respect to the cooling liquid is plotted. Full line: with tungsten plate; broken line: without plate.

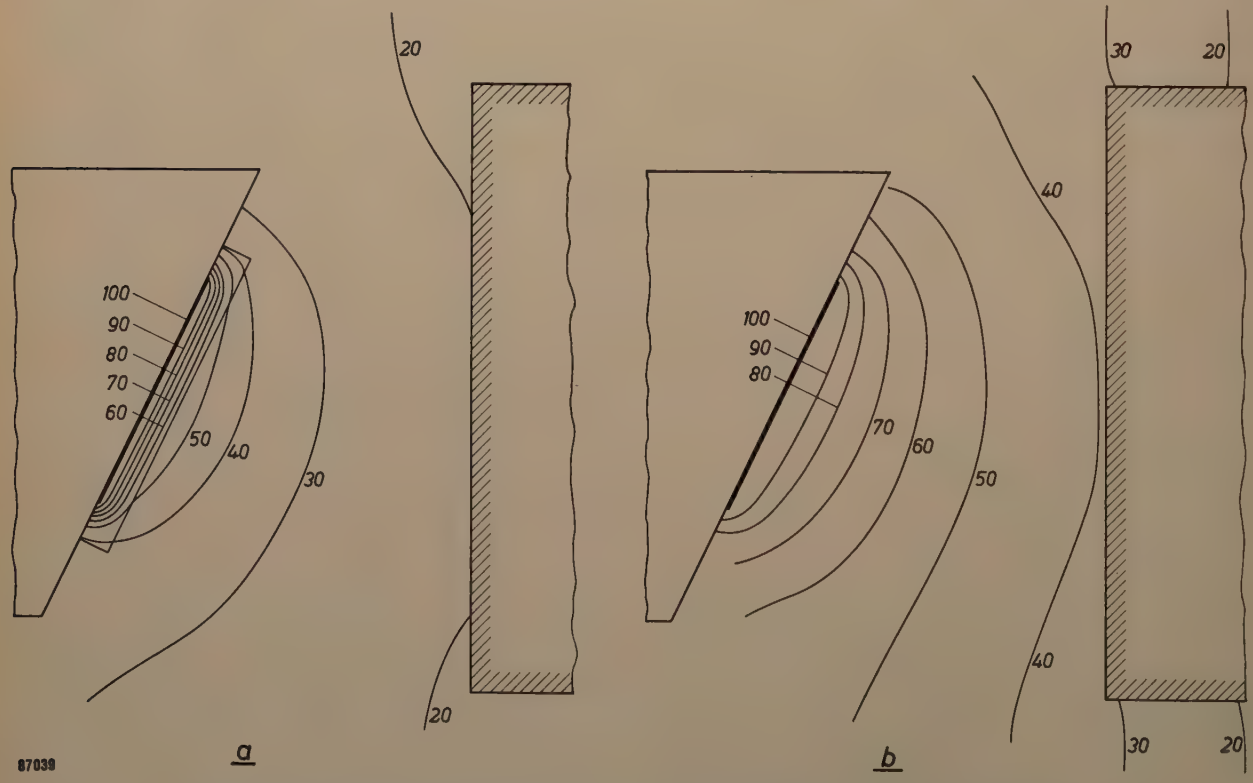


Fig. 11. Equipotential lines measured in a 10 : 1 scale model, *a*) with and *b*) without a tungsten plate using an anode of thickness $A = 25$ mm. The focus dimensions were 9 mm × 20 mm (the same as in fig. 10).

It can be seen from the measurements that the thermal conductivity of the model when the thickness of A is reduced, for example from 25 mm to 15 mm, increases but little. On the other hand, as is shown by fig. 10, the maximum temperature at the oil-cooled surface increases still further by about 10% of the temperature difference between focus and cooling liquid. Consequently, by virtue of there being less risk of the cooling liquid carbonizing with a more uniform temperature distribution of the cooled surface, a thicker anode is to be preferred in practice in spite of the lower value of L that goes with it.

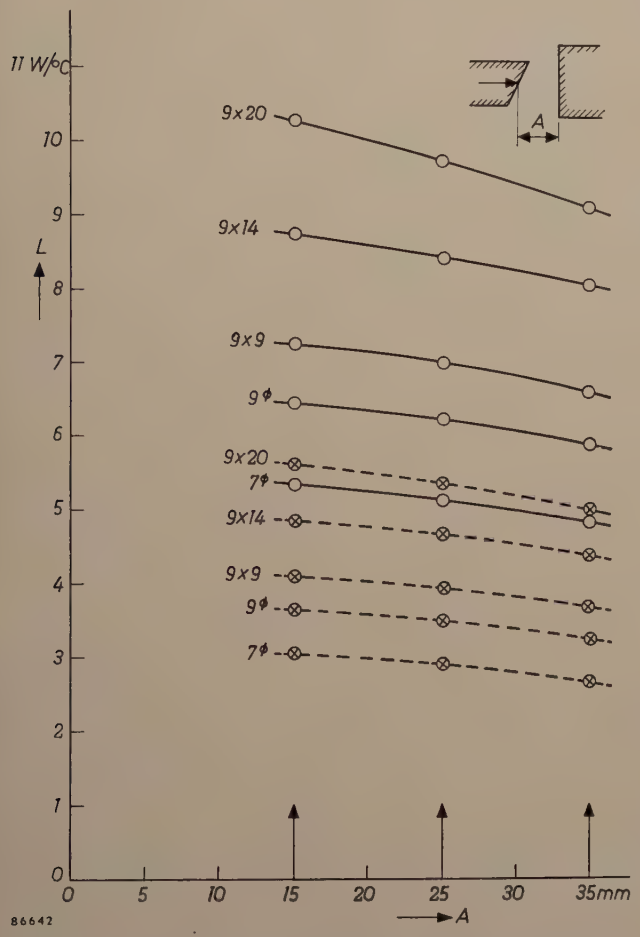


Fig. 13. Thermal conductivity L (reciprocal of the thermal resistance) as a function of the thickness A for various sizes of the focus (dimensions in mm). Full lines: without tungsten plate; broken lines: with plate.

By means of model tests such as described in this article the behaviour of an X-ray anode can be fairly accurately predicted. In particular, the effect of the area of the focus, the thickness of the anode and of the tungsten plate on the mean temperature of the focus can be found from the tests. The thermal phenomena in the anode are clearly reflected by measurements in the electrolytic tank, these measurements providing a valuable supplement to theoretical calculations [7]) and excelling them in clarity.

Appendix: Mathematical basis of the analogies used

In the electrostatic field, in the absence of space charge, the following equation applies [8])

$$\frac{\partial}{\partial x} D_x + \frac{\partial}{\partial y} D_y + \frac{\partial}{\partial z} D_z = 0. \quad . \quad . \quad . \quad . \quad (9)$$

If the medium is isotropic and homogeneous (ε = constant), then in accordance with (1) and after division by ε, we can write instead of (9):

$$\frac{\partial}{\partial x} E_x + \frac{\partial}{\partial y} E_y + \frac{\partial}{\partial z} E_z = 0, \quad . \quad . \quad . \quad . \quad (10)$$

or

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} = 0. \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

If the medium is not homogeneous (ε dependent upon x, y and z), then substituting (10) for (9) is not permissible. If the variation in ε however is discontinuous so that ε has the constant values ε_1 and ε_2 at both sides of a specific boundary, then if D_n represents the component perpendicular to the boundary, we have:

$$(D_n)_1 = (D_n)_2, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

$$\varepsilon_1 (E_n)_1 = \varepsilon_2 (E_n)_2. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

In order that the potential U has the same value at both sides of the boundary, we have, for the tangential component E_t:

$$(E_t)_1 = (E_t)_2. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (14)$$

For the electrical current intensity **j** in a conducting medium, we have, analogous to (9), the equation:

$$\frac{\partial}{\partial x} j_x + \frac{\partial}{\partial y} j_y + \frac{\partial}{\partial z} j_z = 0, \quad . \quad . \quad . \quad . \quad (15)$$

which shows that nowhere does a space charge occur as a result of the current. If the medium is homogeneous (γ = const.), then by reason of (2) we obtain the same equations (10) and (11) as above. The potential accordingly fulfils the same differential equations as in the electrostatic case.

If γ is a function of x, y and z, then substituting (10) for (15) is not permissible. If γ increases from the constant value γ_1 to the constant value γ_2 at a certain boundary, then for the normal component of **j** we have:

$$(j_n)_1 = (j_n)_2, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (16)$$

$$\gamma_1 (E_n)_1 = \gamma_2 (E_n)_2, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (17)$$

whilst the tangential component again fulfils (14) because the potential has to be equal along both sides of the boundary.

---

[7]) See, in this connection, W. J. Oosterkamp, The Heat Dissipation in the Anode of an X-ray Tube, Philips Res. Rep. **3**, 303-317, 1948. In this publication no account is taken of the temperature discontinuity between the anode metal and the cooling liquid, whereas in the present experiments this is an important variable.

[8]) This can be derived from the equation $\int D_n \mathrm{d}A = Q$, which holds for an enclosed surface containing the charge Q. In this, D_n is the component of **D** perpendicular to the surface element $\mathrm{d}A$. If there is no charge then $Q = 0$ and if, this being so, we apply the equation to an element with sides $\mathrm{d}x$, $\mathrm{d}y$ and $\mathrm{d}z$, we obtain equation (9).

From the fact that in both cases we arrive at equation (11), the analogy between the electrostatic field in a homogeneous dielectric and the current field in a homogeneous conductor follows. If the medium in the electrostatic case consists of zones which are individually homogeneous but for which ε has various values ε_1, ε_2 etc., then we can still work with the model provided we ensure that corresponding zones with γ values γ_1, γ_2 etc. are present in the tank It follows from the boundary conditions (13) and (17) that we have to make $\varepsilon_1 = \gamma_1$, $\varepsilon_2 = \gamma_2$ etc.. If we are merely concerned with ratios of potential it is sufficient if $\varepsilon_1 : \gamma_1 = \varepsilon_2 : \gamma_2$ etc.

In the case of constant heat flow, we have the equation:

$$\frac{\partial}{\partial x} q_x + \frac{\partial}{\partial y} q_y + \frac{\partial}{\partial z} q_z = 0, \quad . \quad . \quad . \quad . \quad (18)$$

which shows that nowhere are there any sources or sinks of heat. If λ is independent of x, y and z then on account of (3) we may replace (18) by:

$$\frac{\partial}{\partial x} \tau_x + \frac{\partial}{\partial y} \tau_y + \frac{\partial}{\partial z} \tau_z = 0, \quad . \quad . \quad . \quad (19)$$

or

$$\frac{\partial^2 \vartheta}{\partial x^2} + \frac{\partial^2 \vartheta}{\partial y^2} + \frac{\partial^2 \vartheta}{\partial z^2} = 0. \quad . \quad . \quad . \quad . \quad . \quad . \quad (20)$$

If λ is a function of x, y and z, then substitution of (18) by (19) is not permissible. At a boundary between two media where λ increases from the constant value λ_1 to the constant value λ_2, we have, for the normal component:

$$(q_n)_1 = (q_n)_2, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (21)$$

or

$$\lambda_1 \left(\frac{\partial \vartheta}{\partial n}\right)_1 = \lambda_2 \left(\frac{\partial \vartheta}{\partial n}\right)_2. \quad . \quad . \quad . \quad . \quad . \quad (22)$$

Since the temperature must be equal at both sides of the boundary, we have, for the tangential component:

$$\left(\frac{\partial \vartheta}{\partial t}\right)_1 = \left(\frac{\partial \vartheta}{\partial t}\right)_2. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (23)$$

From equations (20) and (11) can be seen the analogy between the case of the constant heat flow and the electrical current in a homogeneous conducting medium. The potential u corresponds to the temperature difference ϑ. When we have various homogeneous media with different λ values, in order that the model should be a replica of the original we must ensure that $\lambda_1 : \gamma_1 = \lambda_2 : \gamma_2$ etc., as shown by equations (22) and (17).

It must furthermore be stated that γ is substantially independent of u (Ohm's law) in the electrical case, whereas in general λ does depend slightly on T and therefore on ϑ. No account is taken of this fact in the model.

Summary. The analogy existing between the electric field of a potential in an electrolyte and the electrostatic field in a dielectric is mentioned. A similar analogy exists between the potential field in an electrolyte and the temperature field in a conductor under conditions of steady heat flow. Methods of setting up an electrical model (electrolytic tank) for a given thermal conductivity problem are discussed and from this model the required information (temperature distribution in the conductor, thermal resistance) can be obtained. Account must be taken of possible temperature discontinuities e.g. between a cooled metal wall and the cooling liquid. The method is applied to the study of the thermal behaviour of an oil cooled copper X-ray anode carrying an embedded thin plate (pastille) of tungsten. In an appendix, the basis of the analogy of the model is explained by comparing the appropriate differential equations.

BEHAVIOUR OF THE TRANSISTOR AT HIGH CURRENT DENSITIES

by F. H. STIELTJES and L. J. TUMMERS.

537.311.33:621.375.4

The current amplification factor of a transistor, which is of great importance for the power amplification obtainable, is found to be dependent upon the current density inasmuch as the current amplification factor falls off if the current density is raised sufficiently. An extension of the theory given in an earlier article, to include those cases where the minority concentrations are not everywhere negligible with respect to the majority concentrations, will account for this effect and indicate how it can be reduced.

According to the simple theory of the junction transistor as given in an earlier article in this Review [1]), hereafter referred to as I, the current amplification factor α of a transistor does not depend upon the applied emitter voltage V_e (*fig. 1*), since this voltage does not appear in the approximate formula derived for α (I, 13) nor is it found in the more exact expression that can be obtained by the product of formulae (I, 19) and (I, 20). In practice, however, we find that when the current is increased (by raising the emitter voltage) the value of α, after an initial increase, declines to a value far below its initial value valid for low current densities (see *fig. 2*, in which $1-\alpha$ is plotted). The failure of the simple theory is attributable to the fact that one of its premises is no longer tenable when the current density is raised. In fact, as we shall demonstrate, the minority concentration in the base then ceases to be everywhere very small in comparison with the majority concentration at the same spot. This has two important consequences: firstly, the minority current in the base is no longer purely a diffusion current, for the electric field also begins to have a marked effect; secondly, the expression given in I (I, 1) for the minority concentration in the boundary planes of a *P-N* junction is no longer valid. These two facts, which were first pointed out by Webster and by Rittner [2]) will be considered more closely. Subsequently we shall discuss the explanation by these authors of the form of the $1-\alpha$ curve as shown in fig. 2. From this theoretical explanation we shall deduce that the value of the current density at which the drop of α becomes troublesome can be substantially raised by the incorporation of a large number of acceptor atoms in the emitter material. In the Research Laboratories at Eindhoven good results have been obtained by the addition of a small amount of gallium to the indium used in the preparation of the emitter of a *P-N-P* alloy transistor.

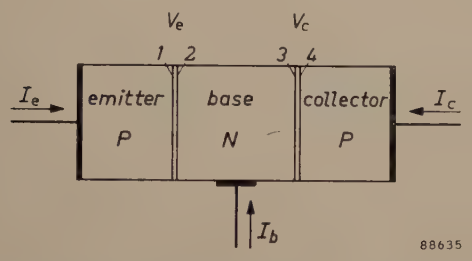


Fig. 1. Schematic representation of a *P-N-P* transistor. I_e, I_c and I_b are respectively the emitter, collector and base currents; *1*, *2*, *3* and *4* mark the boundary planes of the *P-N* and *N-P* junctions. V_e is the voltage between the planes *1* and *2*; V_c is the voltage between the planes *4* and *3* and has a negative value (inverse voltage).

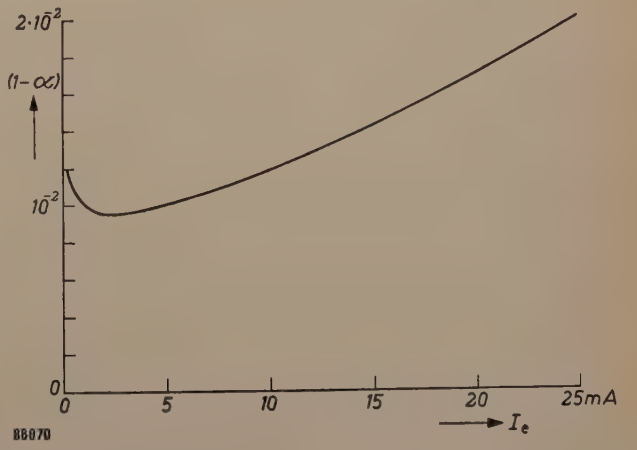


Fig. 2. Only a fraction α of the total emitter current I_e reaches the collector. $1-\alpha$ can, therefore, be considered as a loss. The loss is at a minimum for a relatively low current density, but it climbs continuously as I_e is further increased. This may be seen from the curve above, which was derived from measurements.

[1]) F. H. Stieltjes and L. J. Tummers, Simple theory of the junction transistor Philips tech. Rev. **17**, 233-246, 1955/56 (No. 9). This article may also be referred to for explanations of the usual transistor terminology concerning holes and electrons, *P* and *N* regions, *P-N* junction, majority and minority concentrations, etc.

[2]) W. M. Webster, On the variation of junction-transistor current-amplification factor with emitter current, Proc. I.R.E. **42**, 914-920, 1954; E. S. Rittner, Extension of the theory of the junction transistor, Phys. Rev. **94**, 1161-1171, 1954.

Breakdown of the simple theory at high current densities

Effect of the field on the minority current at high current densities

Consider the emitter junction of a *P-N-P* transistor. The emitter efficiency γ, i.e. the fraction of the emitter current transported by holes, lies close to unity for a good transistor. In the base the current from emitter to collector is therefore almost entirely composed of holes; the electron current is here virtually zero. At every point of the base where the hole concentration (minority concentration) is very small with respect to the electron concentration (majority concentration), that is, at every point where

$$p_N/n_N \ll 1 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

applies [3]), the hole current is almost a pure diffusion current. If the hole concentration is so high that (1) is no longer valid, then the electric field also plays a part in determining the hole current.

The proof of these statements, which in this special case (majority current practically zero) is much easier to grasp than the general proof given in I (page 235, small print), is as follows:

As in the general proof, we start from the assumption that hardly any space charge can arise outside a *P-N* junction (neutrality condition). This implies that the concentration gradients for the two sorts of charge carriers are the same in any cross-section outside a *P-N* junction (*fig. 3*). Thus there is not only a diffusion current of holes, but also one of electrons; indeed, the electron diffusion current is about twice as great as the hole diffusion current, for the diffusion constant for electrons is about twice that for holes. Diffusion causes holes and electrons to move in the same direction and hence the electric current densities, designated by J_{pD} for holes and J_{nD} for electrons, are of opposite sign so that

$$J_{nD} = -2\,J_{pD} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

If, in spite of this, the total electron current remains (virtually) zero, then there must be an electric field that brings about a field current of electrons (with a density J_{nF}) which cancels out the electron diffusion current, i.e. $J_{nF} = -J_{nD}$. Replacing J_{nD} in (2), we can write

$$J_{nF} = 2\,J_{pD}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

If the concentration of electrons is very high compared to that of the holes, the field creates a hole current J_{pF} which is negligible with respect to J_{nF}, and, in accordance with (3), with respect to J_{pD} too; if this condition is not satisfied, however, the field current of the holes cannot be disregarded.

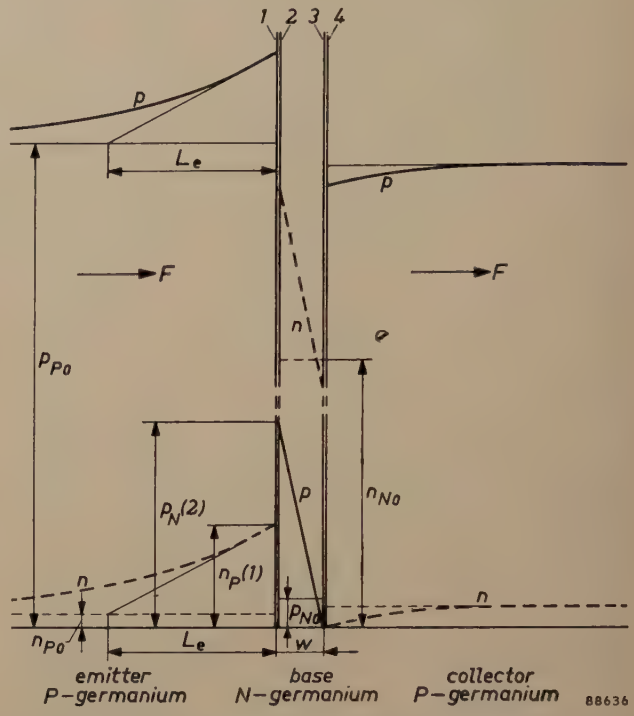


Fig. 3. Hole concentrations (full lines) and electron concentrations (dotted lines) in a *P-N-P* transistor. *1*, *2*, *3* and *4* boundary planes of the *P-N* and the *N-P* junctions. p_{P0} and n_{P0} equilibrium concentrations of holes and electrons in the emitter. p_{N0} and n_{N0} the same in the base. The equilibrium concentrations in the collector are not dealt with in this article and are therefore not represented by symbols. $n_P(1)$ and $p_N(2)$ minority concentrations in the boundary planes *1* and *2*; these are fixed by the voltage applied between *1* and *2*. In the boundary planes *3* and *4* the inverse voltage applied between these planes fixes the minority concentrations at zero. L_e diffusion length in the emitter, w thickness of base. The electric field F acts in the direction of the arrows.

The electric neutrality of the base has the consequence that the difference [4]) between the electron concentration n_N and the hole concentration p_N is the same at any point and is thus equal to the difference between the equilibrium concentrations n_{N0}—p_{N0} (fig. 3). If we neglect p_{N0} in comparison with n_{N0}, we obtain

$$n_N - p_N = n_{N0} \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

This being so, condition (1) is equivalent to

$$p_N/n_{N0} \ll 1,$$

[3]) In this article we shall have to take account of the majority concentrations to a far greater extent than was done in I. We must therefore use a somewhat different notation. Hole and electron concentrations will be indicated by p and n respectively. The suffix P or N indicates whether a magnitude is related to a *P* or to an *N*-region. It is thus immediately apparent whether a minority concentration (p_N and n_P) or a majority concentration (p_P and n_N) is concerned. The corresponding equilibrium concentrations are denoted by an additional suffix 0. The suffixes p and n are used to denote whether a quantity refers to holes or to electrons; e.g. J_p stands for the hole-current density, and D_n for the diffusion constant for electrons.

[4]) One should remember that the nett charge of the *mobile* charge carriers is balanced by the nett charge of the (fixed) donors and/or acceptors present; electric neutrality does not therefore imply that $p_N = n_N$. See eq. (5) in J. C. van Vessem, The theory and construction of germanium diodes, Philips tech. Rev. **16**, 213-224, 1954/55.

since, if either of these conditions is satisfied, it follows from (4) that $n_N \approx n_{N0}$, so that the other condition is also fulfilled.

If this condition is satisfied in cross-section 2, in other words if

$$p_N(2)/n_{N0} \ll 1, \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

it is evidently satisfied elsewhere in the base, as may be seen from fig. 3. Hence, if we assume that (5) is satisfied, we may conclude that the hole current in the base is almost exclusively a diffusion current. Its current density J_p is given by:

$$J_p = qD_p \frac{p_N(2)}{w},$$

where q is the magnitude of the electronic charge, D_p is the diffusion constant for holes and $p_N(2)/w$ is the concentration gradient for the holes (fig. 3). Since the electron current is negligible, J_p is equal to the total current density J. Hence it follows from (5) that

$$J \ll qD_p \frac{n_{N0}}{w}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

If, therefore, the current density becomes so high that (6) is no longer satisfied, then condition (5), from which it was deduced, cannot be fulfilled either, and consequently the minority concentration in the base is not everywhere negligible with respect to the majority concentration.

That this situation can occur in practice becomes evident when we insert in (6) the numerical values applicable to a normal *P-N-P* transistor, chosen as an example. (As we shall make repeated use of this same example in the course of this article we shall give some additional data for later reference, apart from that of immediate concern.)

The equilibrium concentrations in the emitter are, say:

$$p_{P0} \approx 5 \times 10^{18} \text{ cm}^{-3}, \; n_{P0} \approx 10^8 \text{ cm}^{-3},$$

and in the base

$$n_{N0} \approx 5 \times 10^{14} \text{ cm}^{-3}, \; p_{N0} \approx 10^{12} \text{ cm}^{-3}.$$

Let the base thickness be

$$w \approx 50 \; \mu = 5 \times 10^{-3} \text{ cm}.$$

The diffusion constants of holes and electrons are

$$D_p = 44 \text{ cm}^2 \text{ sec}^{-1}, \; D_n = 88 \text{ cm}^2 \text{ sec}^{-1}.$$

The average lifetime of the holes in the base, determined by recombination in the interior and at the surface of the base, is

$$\tau_b \approx 15 \; \mu\text{sec}.$$

The area of the emitter junction is 3×10^{-3} cm^2 (0.3 mm^2). As $q = 1.6 \times 10^{-19}$ coulomb, we find from (6) that for this particular transistor the influence of the field on the hole current is negligible provided that the current density

$$J \ll 0.7 \text{ A/cm}^2,$$

i.e. as long as the emitter current is considerably less than 2 mA. Currents of a few mA, however, are not exceptional, whilst far heavier currents occur with transistors in output-amplifier circuits. In such cases the minority current in the base will no longer be purely a diffusion current, and the influence of the field will have to be taken into account.

Minority concentrations in the boundary planes of a P-N junction

The theory put forward in article I was based on the assumption that the minority concentrations in the boundary planes *1* and *2* of a *P-N* junction are fixed by the applied voltage V at

$$n_p(1) = n_{P0} \, e^{qV/kT} \text{ and } p_N(2) = p_{N0} \, e^{qV/kT}, \quad (7a, b)$$

respectively (see I, (1)).

For the sake of convenience we shall introduce in place of V a new variable Z, defined as:

$$Z = Q \, e^{qV/kT}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

in which

$$Q = \frac{p_{N0}}{p_{P0}} = \frac{n_{P0}}{n_{N0}}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

Q is thus the ratio of the equilibrium minority concentration at one side of the *P-N* junction to the equilibrium majority concentration at the other side, and is therefore a very small number. In the transistor chosen as example, $Q = 2 \times 10^{-7}$. The equivalence of the two expressions for Q follows from the condition $p_{N0}n_{N0} = p_{P0}n_{P0}$ (the product of the equilibrium concentrations of majority and minority charge carriers is independent of the region considered.) Z is evidently always positive; only if $V = -\infty$ will Z become zero. We shall further introduce B, the ratio of the equilibrium majority concentration in the *P*-region to that in the *N*-region:

$$B = \frac{p_{P0}}{n_{N0}}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

In a transistor the equilibrium majority concentration in the emitter always substantially exceeds that in the base (cf. I, page 240), so that B is numerically large. In our sample transistor $B = 10^4$.

(7a) and (7b) can now be written in the more convenient form:

$$n_P(1) = n_{N0}Z, \quad . \quad . \quad . \quad . \quad . \quad (11a)$$

$$p_N(2) = n_{N0}BZ. \quad . \quad . \quad . \quad . \quad (11b)$$

In view of our intention to extend the theory to cover higher current densities, these formulae require correction. The values of the four unknown concentrations, viz. the majority and minority concentrations in the two boundary planes of a *P-N* junction, are given for the general case by four equations which were worked out in an earlier article in this Review [5]). By solving these equations we obtain for the minority concentrations:

$$n_P(1) = n_{N0}Z\,\frac{1+BZ}{1-Z^2}, \quad . \quad . \quad . \quad (12a)$$

$$p_N(2) = n_{N0}Z\,\frac{B+Z}{1-Z^2}, \quad . \quad . \quad . \quad (12b)$$

expressions which, for sufficiently small values of Z, reduce to (11a) and (11b). We see from these that both $n_P(1)$ and $p_N(2)$ would become infinitely great if Z equalled unity.

According to (8), $V = (kT/q)\ln(1/Q)$ if $Z = 1$. Formula (9) of the article referred to in [4]) shows that this value represents the spontaneous potential difference (contact potential) between the *N*-region and the *P*-region. An external voltage compensating the contact potential would thus involve infinitely high minority concentrations and hence an infinitely great current. It will be clear that in practice the external voltages appearing across the *P-N* junction always lie below the contact potential. In our sample transistor we arrive at a value of 0.3 V for the contact potential (at room temperature $kT/q = 1/40$ V). The voltages that are applied across the crystal, may, however, be higher than 0.3 V, since the series resistance offered by the crystal itself causes a considerable potential drop.

Z, therefore, is always less than unity. We shall demonstrate in due course that even at a very high current density the emitter voltage is still so low that $Z \ll 1$. The terms Z^2 in the denominators of the formulae (12a) and (12b) may therefore be disregarded. Moreover, in view of the high value of B, Z may be neglected in comparison with B in the numerator of (12b), but, for the very same reason, it is not necessarily permissible to neglect BZ in comparison with unity in formula (12a), so that

$$n_P(1) = n_{N0}Z\,(1+BZ), \quad . \quad . \quad . \quad (13a)$$

$$p_N(2) = n_{N0}BZ. \quad . \quad . \quad . \quad . \quad . \quad (13b)$$

Equation (13b) is identical with (11b), but (13a) can be replaced by (11a) only if $BZ \ll 1$, which implies, according to (13b), that $p_N(2) \ll n_{N0}$. Here we encounter once more the same condition (see (5)) as the one which makes it possible to treat the minority current in the base as a pure diffusion current. From (13a) and (13b) it follows that, as Z increases from zero (i.e. with increasing voltage in the forward direction and hence increasing current density), the ratio $n_P(1)/p_N(2)$, which in the simple theory was considered to remain constant (viz. $1/B$) also increases. This causes, as we shall presently demonstrate, a loss in emitter efficiency.

The transistor losses as a function of the current density

The current amplification factor α is the product of the base efficiency β and the emitter efficiency γ (cf. I, (9)). For the purpose of this article we shall work in terms of losses, namely the total transistor loss $1-\alpha$, the base loss $1-\beta$, and the emitter loss $1-\gamma$. Because $1-\beta$ and $1-\gamma$, and hence also $1-\alpha$, are in practice always small with respect to unity, it follows from $\alpha = \beta\gamma$ that we may write [6])

$$1-\alpha = (1-\beta) + (1-\gamma). \quad . \quad . \quad . \quad (14)$$

Once we have established the behaviour of $1-\beta$ and $1-\gamma$ as functions of the current density, we shall know that of $1-\alpha$ as a function of the same quantity. Before proceeding to do so, however, we must examine more closely the effect of the electric field on the minority charge carriers in the base of a transistor.

Apparent increase of the diffusion constant of minority charge carriers owing to the influence of the field

Field and diffusion drive the minority carriers in the same direction (fig. 3). The effect of the field may thus be regarded as an apparent increase in the diffusion constant. A fact of great importance is that, for very high minority concentrations, the apparent diffusion constant approaches a limit which is twice the actual diffusion constant. We can explain this as follows. The hole current density J_p

[5]) See the article referred to in footnote [4]), equations (12) to (15), in which a is to be replaced by q/kT and $-\Delta\varphi$ by V. Strictly speaking, in formulae (12a and (12b) of the present article, the factor n_{N0} should be replaced by $n_{N0}-p_{N0}$, whilst B should be written as $\frac{p_{P0}-n_{P0}}{n_{N0}-p_{N0}}$. There is, however, no objection to neglecting p_{N0} and n_{P0} in comparison with n_{N0} and p_{P0} respectively.

[6]) By substituting $\beta = 1-(1-\beta)$ and $\gamma = 1-(1-\gamma)$ in $\alpha = \beta\gamma$, and neglecting the very small term $(1-\beta)(1-\gamma)$

and the electron current density J_n, each written as the sum of the contributions of diffusion and field can be expressed by

$$J_p = -q\, D_p \frac{dp}{dx} + qp\mu_p F \quad . \quad . \quad . \quad (15a)$$

$$J_n = q\, D_n \frac{dn}{dx} + qn\mu_n F. \quad . \quad . \quad . \quad (15b)$$

respectively (μ_p and μ_n representing the mobilities of holes and electrons. Apart from this, we have at our disposal the Einstein relation, which expresses the logically acceptable fact that there exists a direct proportionality between the diffusion constant D and the mobility μ for particles having an electric charge of equal magnitude:

$$\frac{D_p}{\mu_p} = \frac{D_n}{\mu_n} = \frac{kT}{q}. \quad . \quad . \quad . \quad . \quad . \quad (16)$$

Finally we know from the neutrality condition (valid for regions outside a *P-N* junction) that the concentration gradients for holes and electrons are equal:

$$\frac{dn}{dx} = \frac{dp}{dx} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (17)$$

By eliminating F from (15a) and (15b) and using (16) and (17), we obtain for J_p:

$$J_p = -q \frac{1 + \frac{p}{n}}{1 - \frac{J_n}{J_p}\frac{\mu_p}{\mu_n}\frac{p}{n}} D_p \frac{dp}{dx} = -q\, D_p' \frac{dp}{dx}. \quad (18)$$

This expression for J_p has the form applicable to a diffusion current density. The influence of the field is expressed in the substitution of the apparent diffusion constant D_p' for the ordinary diffusion constant D_p. In an N-region, where we write p_N instead of p, we find after substitution from $\mu_p/\mu_N = \frac{1}{2}$ and from $n_N = p_N + n_{N0}$ (neutrality condition, see (4)), that

$$D_p' = \frac{1 + \frac{p_N}{p_N + n_{N0}}}{1 - \frac{J_n}{2J_p}\frac{p_N}{p_N + n_{N0}}} D_p. \quad . \quad . \quad (19a)$$

As the current in the base of a transistor is mainly composed of holes, $J_n/J_p \ll 1$, allowing us to write:

$$D_p' = \left(1 + \frac{p_N}{p_N + n_{N0}}\right) D_p, \quad . \quad . \quad (19b)$$

this changing for very high values of the minority concentration p_N into:

$$D_p' = 2\, D_p. \quad . \quad . \quad . \quad . \quad . \quad . \quad (20)$$

Fig. 4 shows how D_p'/D_p gradually increases from 1 to 2 as the value of p_N/n_{N0} increases. The apparent increase in the diffusion constant is noticeable at high current densities, since these are accompanied by high minority concentrations in the base.

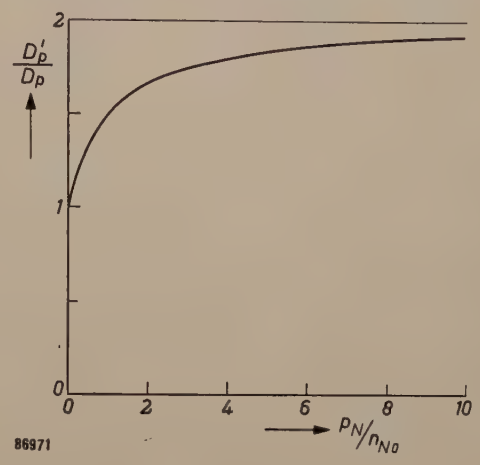


Fig. 4. The ratio D_p'/D_p of the apparent to the actual diffusion constant as a function of the ratio p_N/n_{N0} of the minority concentration to the equilibrium majority concentration in the base of a *P-N-P* transistor.

Reduction of base losses at high current densities

Owing to the effect of the field in the base of a *P-N-P* transistor at high current densities, the holes are transported more quickly through the base and thus run less risk of loss through recombination. The base loss, therefore, will diminish at increasing current density. The hole concentration is always highest near the emitter, in cross-section *2* (fig. 3). Here it will be most noticeable that the apparent diffusion constant D_p' deviates from the normal value D_p and it is here too that D_p' will come nearest to its limit of 2 D_p. At very high current densities (theoretically only at an infinitely high value) D_p' will reach the value 2 D_p throughout the whole base (i.e. up to infinitely near the collector). The average time a hole spends in crossing the base will then be halved. If we assume that the average lifetime τ_b of a hole in the base is independent of current density, then that proportion of the total number of holes arriving per second in the base lost by recombination, will also be halved. Hence, at increasing current density, the base loss approaches half the value it has at low current densities. Upon further analysis (into which we

shall not enter here) it is found that the base loss can be represented to a close approximation by simply substituting D_p' for D_p in the expression (I, 12). If in addition D_p' is replaced by the expression (19b), we find that:

$$1-\beta = \frac{w^2}{2\,\tau_b D_p}\,\frac{1+p_N(2)/n_{N0}}{1+2p_N(2)/n_{N0}} \quad . \quad . \quad (21).$$

In *fig. 5* $1-\beta$ is plotted as a function of $p_N(2)/n_{N0}$, on the basis of the numerical values given in our example.

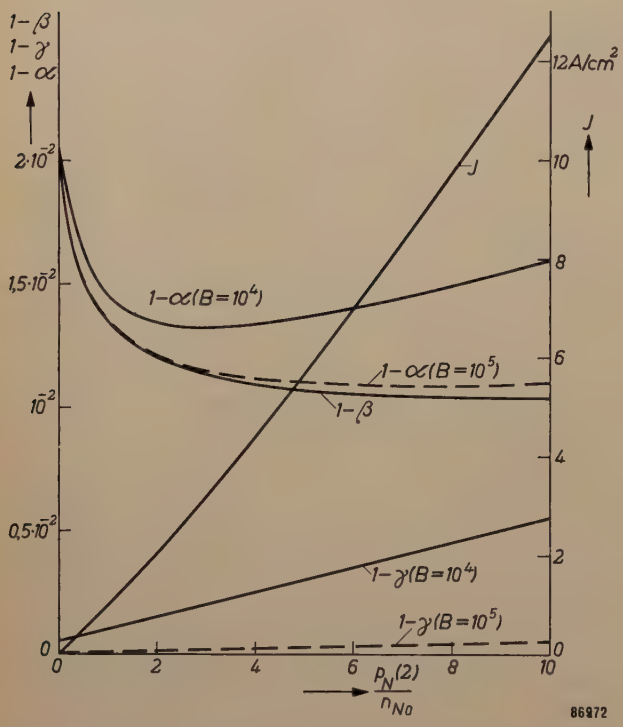


Fig. 5. Base loss $1-\beta$ (β = base efficiency), emitter loss $1-\gamma$ (γ = emitter efficiency) and current density J as functions of $p_N(2)$ (minority concentration in the boundary plane of the base at the emitter side), using n_{N0}, the equilibrium majority concentration in the base, as the unit. The full curves apply to the transistor chosen as an example in this article, for which $B\ (= p_{P0}/n_{N0}) = 10^4$; the dotted curves refer to a transistor whose equilibrium concentrations in base and emitter have been so modified that $B = 10^5$.

The premise that τ_b, the average lifetime of the holes in the base, is independent of the current density, i.e. independent of the concentrations of holes and electrons, even when the concentrations are very high, is not an immediately obvious one. In such circumstances every hole encounters a very large number of electrons, and one would expect the chance of recombination to increase for every hole, with the consequence that τ_b is shortened. In fact, however, recombination is a very intricate process. Direct recombinations occur very rarely; as a rule the process requires a number of intermediate steps, so that τ_b may well remain virtually constant. The assumption that it does so is supported by the results of measurements.

In the base of a *P-N-P* transistor the hole concentration gradually decreases towards zero from emitter to collector, so that according to (19b) the (apparent) diffusion constant D_p' gradually decreases towards D_p. Disregarding generation and recombination in the base, we may consider J_p to be independent of the position x, and according to (18), dp/dx must then gradually increase from emitter towards collector. The concentration curve of the holes, therefore, is no longer linear. The calculations of base and emitter efficiency in I, however, were based on a supposedly nearly linear concentration curve. This concentration curve can also be calculated at high current densities by integration of (18), after substituting the expression (19b) for D_p'. The boundary conditions are that the values of p at emitter and collector must be $p_N(2)$ and 0 respectively. We obtain in this way the curves shown in *fig. 6*. For very small, but equally so for very large values of $p_N(2)$, the deviation from linearity is negligible. In the latter case p is very large up to the direct vicinity of the collector, so that over almost the whole of the base D_p' has nearly the constant value of $2\,D_p$. Hence dp/dx is also constant, according to (18). Only in the immediate vicinity of the collector will the gradient increase, doubling its value at the collector itself (curve *d*). A maximum deviation from linearity must then occur between the extreme cases of very low and infinitely high values of $p_N(2)$. That this maximum deviation is still fairly small is apparent from the form of curve c. In our example this case occurs at a current of about 8 mA. The approximate formulae (21) and (22) for the base and emitter losses are not affected by this slight deviation from linearity.

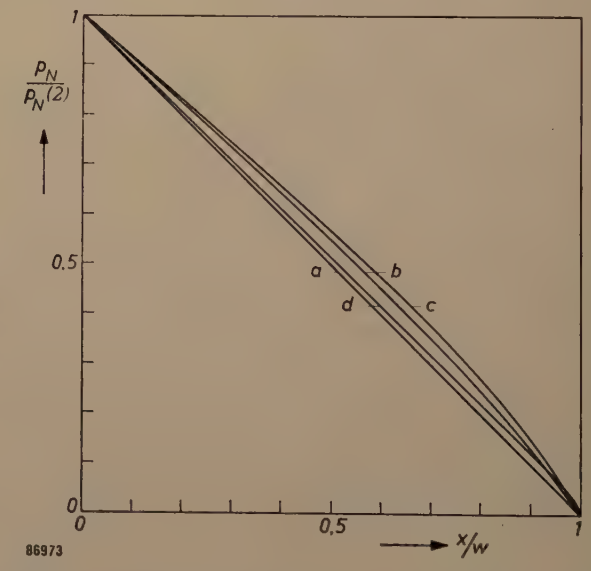


Fig. 6. The hole concentration p_N in the base of a *P-N-P* transistor shows a spatial variation that is always nearly linear. The curves shown here are plotted for different values of $p_N(2)/n_{N0}$. The origin, $x = 0$ lies at cross-section *2*, at the emitter. To facilitate comparison of the curves as regards their linearity, the quantity actually plotted is the ratio $p_N/p_N(2)$, whereby all curves start in point 1 on the ordinate. Curve *a*: $p_N(2)/n_{N0} \ll 1$; curve *b*: $p_N(2)/n_{N0} \approx 1$; curve *c*: $p_N(2)/n_{N0} \approx 2$ (greatest deviation from linearity). For still greater values of $p_N(2)/n_{N0}$ (curve *d*) the curve approaches *a* again.

The increase in the emitter loss at high current densities

We should prefer the total current through the emitter junction to be carried entirely by holes, but in fact a minor part of it is composed of electrons. In the emitter the electron current is a pure diffusion current, so that J_n is proportional to $n_P(1)$ [7]). If the hole current in the base were also a pure diffusion current, then J_p would be proportional to $p_N(2)$ and J_n/J_p would be proportional to $n_P(1)/n_N(2)$ in consequence. However, as we have shown on page 64, this ratio becomes higher as J increases, so that an increasing portion of J would be composed of electrons. There is a certain amount of compensation owing to the fact, likewise discussed earlier, that the electric field comes to the aid of the hole current, so that initially J_p increases (proportionately) more rapidly than $p_N(2)$. Upon further analysis, however, it turns out that this effect of the field cannot fully compensate the effect of the increase of $n_P(1)/p_N(2)$, so that J_n/J_p nevertheless always increases with J. The increase is accompanied by a falling-off of the emitter efficiency γ and hence with an increase in emitter loss $1-\gamma$. The analysis shows that the expression for $1-\gamma$, which could be derived from (I, 11), requires a correction factor $(1+p_N(2)/n_{N0})$. We then arrive at

$$1-\gamma = \frac{D_n}{D_p}\frac{w}{L_e}\frac{1}{B}\left(1+\frac{p_n(2)}{n_{N0}}\right), \quad \ldots \quad (22)$$

in which B again is given by (10). L_e represents the diffusion length in the emitter (fig. 3). The fully drawn curve in fig. 5 applies to $1-\gamma$ plotted for our example [8]).

The total loss $1-\alpha$, consisting according to (14) of the sum of emitter and base losses, is also plotted in fig. 5 as a function of $p_N(2)/n_{N0}$. This diagram further shows the current density $J(=J_p)$, and from this we may derive a curve showing $1-\alpha$ as a function of J. The full curve of *fig.* 7 was obtained in this purely theoretical manner; it does in fact have the same character as the experimentally derived curve of fig. 2.

It appears from the curve for J in fig. 5 that J is already very large when $p_N(2)/n_{N0} = 10$. It follows from (13b) that in this case $BZ = 10$ so that, with $B = 10^4$, $Z = 10^{-3}$. Neglecting Z^2 with respect to 1 and Z with respect to B, as was done in the derivation of (13a) and (13b), was, therefore, perfectly justified.

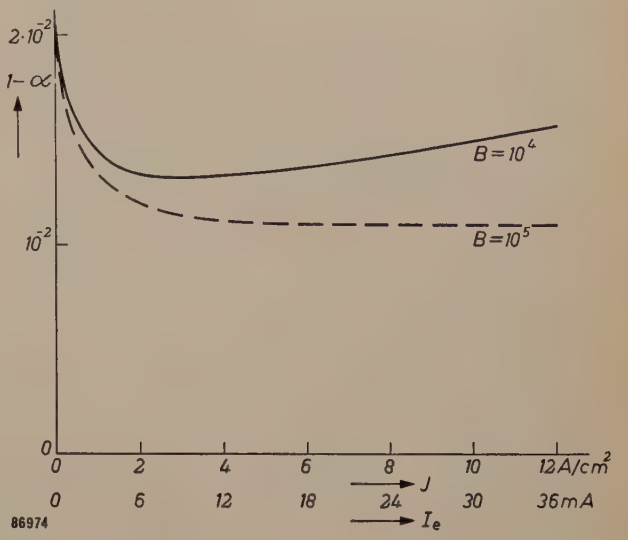


Fig. 7. The losses $1-\alpha$ of a transistor as a function of the current density J, as derived from fig. 5, for two values of B. A scale of emitter current values (I_e) based on an emitter surface of 0·3 mm², is also set out along the abscissa.

Reduction of transistor losses

Formulae (21) and (22) show how various quantities influence the transistor losses. The influence of the factor B, which is defined by (10), is of particular importance. The base loss is evidently independent of B, whereas $1-\gamma$ is inversely proportional to it. If, for example, B is multiplied by a factor of 10, the slope of the curve for $1-\gamma$ will become 10 times smaller. Only for much larger values of $p_N(2)$, then, will $1-\alpha$ (dotted curve in fig. 5) diverge perceptibly from $1-\beta$. Analysis shows that for J we can write:

$$J = q\frac{D_p n_{N0}}{w}\left\{2\frac{p_N(2)}{n_{N0}} - \ln\left(1+\frac{p_n(2)}{n_{N0}}\right)\right\},$$

an expression in which B does not appear, so that the curve J in fig. 5 is independent of the value of B. Thanks to the larger value of B, the curves $1-\gamma$ and $1-\alpha$ as functions of J have the much more favourable form shown as a dotted curve in fig. 7.

In order to raise the value of B $(=p_{P0}/n_{N0})$, we may attempt to increase p_{P0}, the concentration of the holes in the emitter. This can be done by incorporating a large number of acceptor atoms in the germanium of the emitter [9]). For the provision of acceptor atoms, indium is often used, a substance

[7]) See I, p. 236, eq. 4. Strictly speaking, J_n is proportional to $n_P(1) - n_{P0}$, but n_{P0} can in practice always be neglected in comparison with $n_P(1)$.

[8]) The concentration change of the electrons in the emitter as shown in fig. 3 is not quite as it is in fact. The thickness d_e of the emitter is very small (e.g. 20 μ). At the point of contact with the electrode, equilibrium concentrations occur. Owing to this, the concentration gradient of the electrons in cross-section 1 is not determined by L_e, but by d_e. Consequently, we have put $L_e = 20$ μ in the evaluation of formula (22).

[9]) See the article referred to in footnote [4]), especially fig. 2.

which offers certain advantages in the manufacturing of "alloy transistors". As regards its solubility in germanium, however, another acceptor element such as gallium seems preferable, in view of the fact that the highest p_{P0} attainable with indium is approximately 5×10^{18} cm^{-3}; whereas with gallium which is much more soluble in germanium, a concentration ten times higher can be attained. As against this, gallium has certain properties which render it unsuitable as a direct substitute for indium. In our Eindhoven laboratories experiments have been carried out using an acceptor material consisting of indium in which some gallium had been dissolved. In this way, gallium can be dissolved in the germanium together with the indium with the result that p_{P0}, and hence the factor B, can be made 10 times greater. The advantages of indium from the manufacturing point of view are thus combined with the superior solubility of gallium in germanium.

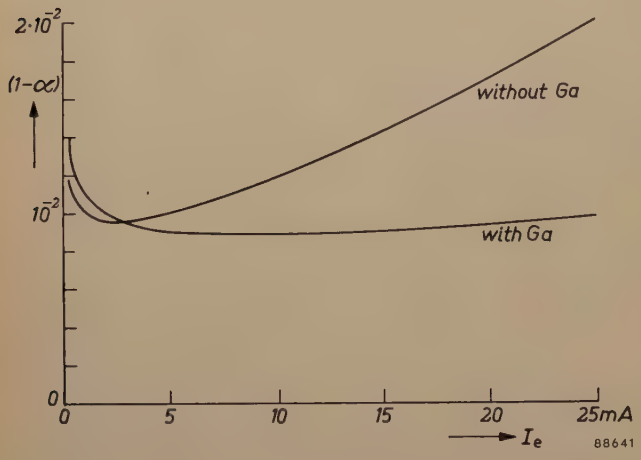


Fig. 8. The transistor losses as a function of the emitter current for a transistor whose emitter contains gallium in addition to indium. For comparison the curve from fig. 2 is also shown, which refers to a similar transistor containing no gallium.

Favourable results have indeed been achieved by this process, as appears from *fig. 8*, in which $1-\alpha$ is plotted as a function of the emitter current I_e in an experimental transistor built on this principle.

A higher equilibrium concentration of majority charge carriers in the emitter is a particularly attractive feature in power transistors (transistors with a relatively large power-handling capacity), in which current densities are always high, but the process can also be profitably applied in the manufacture of transistors not intended for high current densities. Raising p_{P0} allows n_{N0} to be raised while keeping the same values for B and hence for $1-\gamma$ (see (23)). This means that a base material of higher conductivity can be chosen (since the conductivity is higher according as the majority equilibrium concentration — and hence n_{N0} in a base of N material — is higher). This is of advantage for limiting the influence of the "Early effect" (discussed in I, page 245). The same principle can also be applied in the manufacture of transistors for high frequencies, which will be discussed in a subsequent article.

Summary. At high current densities the simple theory of the junction transistor is no longer valid, in view of the fact that in the base the minority concentration ceases to be everywhere negligible with respect to the majority concentration. One consequence of this is that the base loss drops fairly quickly to half its initial value as the current density is raised; another consequence is that the emitter loss gradually increases. The increase in emitter loss continues long after the base loss has become constant, the nett result being that the total losses go on increasing and that the current amplification factor falls off. The higher the equilibrium concentration of majority charge carriers in the emitter, the slower the rise of emitter losses with current density; as a result of this the excessive losses are not incurred until higher current densities. Where *P-N-P* alloy transistors are concerned, this higher concentration can be obtained by the employment of a gallium-indium alloy as acceptor material for the emitter instead of the usual pure indium.

ABSTRACTS OF RECENT SCIENTIFIC PUPLICATIONS BY THE STAFF OF N.V. PHILIPS' GLOEILAMPENFABRIEKEN

Reprints of these papers not marked with an asterisk * can be obtained free of charge upon application to Philips Electrical Ltd., Century House, Shaftesbury Avenue, London W.C. 2.

2281: J. L. Meijering: On the diffusion of oxygen through solid iron (Acta Met. **3**, 157-162, 1955, No. 2).

Internal oxidation of three iron alloys showed — at least at 1200 °C — no trace of intergranular preference. Therefore, in contrast to a conjecture of Kitchener et al., normal interstitial diffusion of oxygen through the iron lattice must be assumed. Approximate values of the permeability (product of diffusion coefficient and solubility) of O in Fe at different temperatures are calculated from the sub-scale thickness. The permeabilities of Fe, Ni, Cu and Ag for oxygen at 85% of their absolute melting temperatures are of the same order of magnitude. Discrepancies in the solubilities of O in solid Fe found by different authors are discussed.

2282: C. J. Bakker: Production and physical properties of radio-isotopes (Rev. Trav. chim. Pays-Bas **74**, 281-293, 1955, No. 5).

The production of radio-isotopes in nuclear reactors and in cyclotrons is discussed. The types

of nuclear reactions involved are mentioned. As to the production of radio-isotopes, the cyclotron is an important complement to the nuclear reactor. Examples of disintegration schemes of radio-isotopes are given. The loss of energy of β or γ-rays during their passage through matter is briefiy discussed. The precautions necessary when working with radio-isotopes are mentioned.

2283: A. M. J. Jaspers and H. R. Marcuse: Local measurement of I^{131} uptake in the thyroid gland (Rec. Trav. chim. Pays-Bas **74**, 355-361, 1955, No. 5).

Based on a new development of Geiger-Müller tubes, a collar-shaped device is described, which can be applied round the patient's neck in order to measure directly the uptake of I^{131} in the thyroid gland. Methods are indicated for calibrating and for measuring the corrections due to the background and the activity in the blood.

2284: J. A. M. Dikhoff and N. W. H. Addink: Ergebnisse von Intensitätsbestimmungen mit der s.p.d. Skala (Microchim. Acta, pp. 257-264, 1955, No. 2-3). (Results of intensity measurements using the standard paper density scale; in German).

Addink has described a simple method for the visual determination of line intensities by means of the so-called standard paper density scale. In the present paper closer attention is given to the background correction, the calibration of the scale, the accuracy of the method and some other factors which may influence the measurements. Finally a short discussion is given of scale lines having a profile similar to that of spectrum lines.

2285: N. W. H. Addink: Leitprobenfreie Verfahren (Microchim. Acta, pp. 703-708, 1955, No. 2-3). (Methods of spectrochemical analysis without standard samples; in German).

It is shown that in the addition method of spectrochemical analysis (in which the unknown concentration of an element is obtained by extrapolation) it is still necessary to check that the intensity-concentration curve is linear when plotted with logarithmic coordinates. Harvey's method is considered, in which the concentration is determined from the ratio of the intensity of a spectral line to the background. In this connection, the intensity of the background may not be used as an "internal standard". Finally details are tabulated of methods of analysis which do not use comparison samples.

2286: N. W. H. Addink and L. J. P. Frank: Zinc in relation to cancer (Naturwiss. **42**, 419-420, 1955, No. 14).

It is demonstrated that fresh venous blood of cancer patients shows a lower zinc concentration than that of healthy people. As, moreover, the zinc content of tumours is lower (about one third) than that of healthy tissues, the zinc content of the blood provides a picture of the tumour.

Examples are given showing favourable and unfavourable progress of the disorder. In the favourable cases the zinc concentration of the blood was gradually increasing with time, whereas in the unfavourable (fatal) ones low values were found. A conjecture concerning the natural and artificial development of the disease is mentioned.

2287: H. C. Hamaker: Naar efficiënte experimenten (Statistica neerlandica **9**, 7-25, 1955, No. 1-2). (Efficient experimentation; in Dutch).

Under industrial conditions which involve large random fluctuations, large numbers of observations and many influencing factors, analysis of data on a statistical basis has become an essential part of experimental techniques, which ultimately leads to more efficient experimentation. Seven examples are given. The first four are simple experiments for the comparison of two batches when disturbing factors such as a time-trend must be eliminated. The other three examples discuss more complex cases, viz. a 3-factor experiment in a foundry, a latin square experiment in nylon spinning and a five-factor experiment of a mixed kind on gear-noise testing. An attempt is made to show in a simple form the relation between the statistical analysis and the conclusions that may be drawn from the results of a well-designed experiment.

2288: G. Diemer: Electroluminescentie (Ned. T. Natuurkunde **21**, 165-176, 1955, No. 7). (Electroluminescence; in Dutch).

Survey of existing theory and experiments on electroluminescence, including the theories of Piper and Williams, of Alfrey & Taylor and of Zalm, Diemer and Klasens. Possible applications of the phenomenon are mentioned.

2289: J. D. Fast and M. B. Verrijp: Solubility of nitrogen in alpha-iron (J. Iron Steel Inst. **180**, 337-343, 1955).

The solubility of nitrogen in α-iron as a function of temperature can be determined accurately by measuring the internal friction of nitrogen-containing iron wires. The value of the internal friction is

proportional to the amount of dissolved nitrogen right up to saturation ($\sim 0.1\%$ by weight). Three different solubilities can be distinguished, viz. those in which the iron is in equilibrium with "Fe_8N", with Fe_4N and with N_2 at atmospheric pressure. All three solubilities were determined. "Fe_8N" is a metastable nitride which forms only below 250 °C from supersaturated solutions. Both "Fe_8N" and Fe_4N form only because of the slowness of the reaction at the outer surface in which N_2 molecules are formed from dissolved N atoms. If this reaction were not slow, the equilibrium between α-iron and "Fe_8N" and Fe_4N could be studied only under large N_2 pressures. These pressures, corresponding to the dissociation pressures of the nitrides can be calculated from the measured solubilities. The results agree satisfactorily with the measurements of American investigators.

2290: H. B. G. Casimir: Ferro- und Antiferromagnetismus (Physikertagung Hamburg 1954, Hauptvorträge, 139-152). (Ferromagnetism and anti-ferromagnetism; in German).

Survey, from a theoretical standpoint, of various aspects of ferromagnetism and anti-ferromagnetism for physicists and electrical engineers not specialized in this field. Two groups of phenomenon are distinguished: those involving the magnetization within a single Weiss-domain and those involving the combined magnetization over a large number of domains. Of the first group, the following are dealt with: saturation magnetization, Curie temperature, the internal magnetic field which results from the exchange force theory of Heisenberg, crystal and strain anisotropy, gyromagnetic resonance and the introduction by Néel of "antiferromagnetism" to explain certain phenomena in ferrites. Néel's view that the magnetization in ferrites is due to two opposed and incompletely compensated magnetized systems (uncompensated antiferromagnetism) is confirmed by experimental work by Gorter at Philips Research Laboratories, Eindhoven.

The phenomena of the second group consists of macroscopic magnetization by the displacement of Block walls and by rotation of the magnetization within Weiss domains. In soft-magnetic alloys, wall displacements are at first responsible for the magnetization — and thus determine the initial permeability — while rotation processes appear only at higher field strengths. In ferrites, there are strong indications that the situation is precisely the reverse. This view is supported by measurements of the A.C. magnetization of ferrites as a function of amplitude and frequency.

The article is concluded with some remarks on hard-magnetic materials.

2291: J. S. C. Wessels: Photo-oxidation of ascorbic acid by isolated chloroplasts (Rec. Trav. chim. Pays-Bas, **74**, 832-840, 1955, No. 7).

The photo-oxidation of ascorbic acid by chloroplasts in the presence of various quinones and dyes was investigated. The kinetics of the reactions were studied, using 2,6-dichlorophenol-indophenol as dye mediator. Inhibitors of the Hill reaction did not affect the photo-oxidation rate. It is postulated that the reaction is photosensitized by chlorophyll. Evidence for this assumption was obtained by bringing about quite the same reaction in alcoholic solution, the chloroplasts being replaced by chlorophyl a. A possible role of this photo-oxidation in the back-oxidation of reduced Hill-reaction oxidants is discussed.

2292: D. J. Braak: New mobile and auxiliary equipment produced by Philips Telecommunication Industries (T. Ned. Radiogenootschap **20**, 161-174, 1955, No. 3).

A new portable trans-receiver has been developed with five channels. Two types are normally made, viz. for the 80 Mc/s band and for the 160 Mc/s band but a third is also made to order for the 40 Ms/c band. After a general description of the equipment and its construction some of the more important auxiliary equipment is briefly described.

2293: L. R. Bourgonjon: Circuits for frequency generators (T. Ned. Radiogenootschap **20**, 175-181, 1955, No. 3).

Present day intensive use of radio equipment calls for higher demands on frequency stability. When high stability must be combined with variation in frequency, it is necessary to use a system for frequency synthesis. A survey of two different systems for frequency synthesis is given, one system using a mixer stage, the other employing a control loop with phase discriminator. To economize on crystals, in both systems an impulse governed oscillator may be used. The block diagram of a multichannel transmitter-receiver in the 100-156 Mc/s frequency band is shown.

2294: H. C. Bennebroek Evertsz: Radio link operating in 4000 Mc/s band for television and multi-channel telephony (T. Ned. Radiogenootschap **20**, 183-192, 1955, No. 3).

A radio link equipment of new design is described. It comprises heterodyne repeater stations and is equipped with S.H.F. triode amplifiers.

2295: W. L. Verwest: Automatic tuning of transmitters (T. Ned. Radiogenootschap **20**, 195-206, 1955, No. 3).

The Philips Instantuner is a device for automatically resetting a tuning element in any one of twelve preset positions. When several tuning elements are used, which is normally the case, a corresponding number of Instantuners is required. Remote control of radio equipment is then limited to a simple manipulation. Manual tuning is also possible in any position by pressing a push-button mounted in a special vernier knob. The presetting procedure is very easy. Changing from one preset frequency to another (automatic tuning) is accomplished in 1-3 seconds for small transmitters, e.g. aircraft transmitters, and in 2-10 seconds for larger transmitters, e.g. broadcast transmitters. Each Instantuner is built as a separate unit, designed for universal application. In consists of a blocking-mechanism and a torque-limiting clutch. The blocking-mechanism contains twelve pawls and pawl rings, so that the main shaft can be blocked in any one of twelve preset positions. Owing to the use of the special torque-limiting clutch a resetting accuracy of $\pm$ 0.01° can be achieved. The clutch releases the driven spindle of a tuning element from the driving-motor, as soon as the preset position is reached.

R 281: W. L. Wanmaker and M. L. Verheyke: A thermogravimetric study on the preparation of calcium halophosphate (Philips Res. Rep. **11**, 1-18, 1956, No. 1).

A thermogravimetric analysis is carried out on the raw materials used in the preparation of the calcium halophosphate phosphor activated by Sb and Mn. Furthermore, some binary and ternary mixtures of these raw materials and some mixtures producing halophosphate on firing are investigated. The dissociation temperature of $CaCO_3$ is depressed by the presence of $CaHPO_4$ and $CaHPO_4 + SrCl_2$. The dissociation of $CaCO_3$ is facilitated to a still greater degree in mixtures producing halophosphates on firing. The dissociation temperature of $CaCO_3$ rises in mixtures producing fluorochloro-, chloro- and fluoroapatite (in the order given). The analysis of halophosphate firing mixtures shows that the loss in Sb, Cl and F occurs only after complete dissociation of the $CaCO_3$. The loss of these elements is high with a low M/6 PO_4 ratio (M = Ca + Sr + Mn). The loss in chlorine may occur as a result of pyrohydrolysis of $SrCl_2$ or of the formation of $SbCl_3$. It is quite probable that the thermobalance could also be used with success for the study of other solid-state reactions in which a loss in weight occurs.

R 282: J. van Slooten: Oscillations and noise (Philips Res. Rep. **11**, 19-26, 1956, No. 1).

In the first section it is shown that there is no well-defined boundary between "order" and "disoder". From a disorderly distribution in time an orderly one may be derived by mathematical or technical means. It is not surprising, therefore, that oscillations and noise may be considered from a common point of view. In the second section the similarity and the difference between a narrow noise band and an oscillator are treated in some detail. Subsequently the phase and frequency fluctuations of an LC oscillator are studied by means of the method of phase jumps.

R 283: J. van den Boomgaard: Zone-melting processes for pure compounds AB with a negligible vapour pressure (Philips Res. Rep. **11**, 27-44, 1956, No. 1).

The composition of a pure compound AB generally will change if zone-melting processes are applied to it. In the present paper the theory for this case is developed under the assumption that reactions with the vapour may be neglected, i.e. the compound has no measurable vapour pressure. With a few general assumptions, which will be clarified by a model describing the solid and the liquid state of binary semiconductors in terms of vacancies, it is possible to define the distribution equilibria with two distribution constants and to derive mathematical expressions for the deviation from the stoichiometric compositions as a function of the place in the rod.

R 284: A. H. Boerdijk: Technical aspects of levitation (Philips Res. Rep. **11**, 45-56, 1956, No. 1).

Levitation of a body is here defined as a state of either stable or neutral equilibrium relative to the earth, in which material contact between the body and its environment is not essential. The possibilities and limitations of levitation by auxiliary gravitational forces, by reaction forces and by forces in electromagnetic fields are investigated. Levitation by gravitational forces or by radiation pressure is not feasible in practice, whilst levitation by forces in electrostatic fields is theoretically impossible. Under certain conditions levitation may be achieved by reaction forces and by forces in magnetostatic, stationary and quasi-stationary electromagnetic fields. Published applications comprise balances

centrifuges, and a method for melting metals in vacuo without a crucible.

R 285: K. F. Niessen: Condition for resonance in a nearly compensated ferrimagnetic (Philips Res. Rep. **11**, 57-65, 1956, No. 1).

The resonance condition is derived for a ferrimagnetic with two sub-lattices with different total moments, different anisotropy constants and different gyromagnetic ratios. These three differences are assumed to be relatively small. The above result may be applied, for example, to an antiferromagnetic in whose sub-lattices a relatively small number of magnetic ions has been replaced by non-magnetic ones.

R 286: B. van der Veen: The equivalent network of a piezoelectric crystal with divided electrodes (Philips Res. Rep. **11**, 66-79, 1956, No. 1).

A piezoelectric-crystal plate with divided electrodes can be considered as a four-pole. A theoretical derivation is given of the equivalent electrical circuit of this four-pole for frequencies close to the mechanical resonant frequency of the plate. The four-pole is compared with the equivalent circuit as put forward by Mason, Sykes, Herzog, and others.

R 287: J. Volger and J. M. Stevels: Electric polarizability of colour centres in quartz crystals and glasses (Philips Res. Rep. **11**, 79-80, 1956, No. 1).

High energy radiation can create colour centres in quartz which give rise to dielectric losses (see these abstracts **R 272**). A possible mechanism to explain these losses is suggested and discussed in connection with various models for colour centres put forward by a number of workers.

R 288: A. van Weel: Error sources in group-delay measurements on electric networks (Philips Res. Rep. **11**, 81-90, 1956, No. 2).

Errors in group-delay measurements can be caused by spurious phase modulation in the amplitude modulator or in the network under test, by a varying carrier-frequency level on the detector, and by overloading the network under test. These effects are discussed and counter measures are indicated.

R 289: J. van den Boomgaard: Zone-melting processes for compounds AB with a measurable vapour pressure under influence of the atmosphere (Philips Res. Rep. **11**, 91-102, 1956, No. 2).

Zone melting of compounds AB, in which deviations from the exact stoichiometric composition may occur and which have a measurable vapour pressure at the melting point, is possible if a vapour of one of the components is applied to prevent the decomposition of the material. If the liquid AB is described in a semi-crystalline way with two kinds of coupled defects, which are in equilibrium with two kinds of coupled defects in the solid, it is possible to describe the distribution equilibrium and the reaction with the gas phase in terms of these defects. With this model, it is possible to derive some mathematical expressions for the relation between the deviation from the exact stoichiometric composition and the position in the rod of AB, as a result of zone-melting processes.

R 290: J. M. Stevels and A. Kats: The systematics of imperfections in silicon-oxygen networks (Philips Res. Rep. **11**, 103-114, 1956, No. 2).

The structure of the SiO_2 network permits a wide variety of possible imperfections, a number of which are discussed. A system is developed by which these imperfections may be described. The system is then extended so as to allow of the description of the structure of glasses. Besides this "synthetic" system for the nomenclature of the centres, an "analytical" system is developed. The advantages of both systems are discussed.

R 291: A. Kats and J. M. Stevels: The effect of U.V. X-ray radiation on silicate glasses, fused silica and quartz crystals. (Philips Res. Rep. **11**, 115-156, 1956, No. 2).

Glasses, fused silica and quartz crystals are compounds with open structures which favour the occurrence of interstitial cations and anions. A systematic investigation is made of the effect of ultraviolet radiation, X-rays and electrons. The centres thus formed are discussed in terms of the new nomenclature developed recently for open structures. Generally speaking, glasses and quartz crystals behave similarly, whereas fused silica has many features in common with amethyst.